

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:50:49 / Search time 60 Seconds  
(without alignment)  
2670.074 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRLTWISFIVGALDSD.....QFDMVPLVILRLRSVTCDD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2993	100.0	567	5	AAE24846	Aae24846 KSHV late
2	125	4.2	777	4	ABE0323	AbB0323 Drosophil
3	122	4.1	364	3	AAE53186	Aab53186 Macaca mu
4	121.5	4.1	2518	3	AAE40574	Aab40574 Human ORF
5	120	4.0	253	3	AAE53182	Aab53182 Macaca mu
6	117	3.9	1316	4	ABG22997	Abg22997 Novel hum
7	115	3.8	1007	4	AAE78838	Aam78838 Human pro
8	115	3.8	2507	6	AEU61812	Abu61812 Human nuc
9	114.5	3.8	1321	4	ABG25416	Abg25416 Novel hum
10	113.5	3.8	1005	7	ADD43960	Add43960 Chlamydia
11	113	3.8	1007	6	ABP96241	Abp96241 Human nuc
12	113	3.8	1007	7	ADC06835	Adc06835 Human pro
13	113	3.8	1019	4	AAE79822	Aam79822 Human pro
14	112.5	3.8	466	7	ADE5522	Ade5522 Rat Prote
15	112.5	3.8	466	7	ADE5524	Ade5524 Human pro
16	112	3.7	1007	7	ADC06800	Adc06800 Human pro
17	111	3.7	509	6	ABU34009	Abu34009 Protein e
18	110.5	3.7	434	3	AAE50910	Aag50910 Arabidops
19	110.5	3.7	975	4	ABE63940	AbB63940 Drosophil
20	108.5	3.6	503	7	AAE58665	Ade58665 Human pro
21	108.5	3.6	503	7	AAE58669	Ade58669 Human pro
22	108.5	3.6	550	4	AAE94880	AbB94880 Human pro
23	108.5	3.6	561	4	AAE40771	Aam40771 Human pol
24	108	3.6	953	6	ABU36789	Abu36789 Protein e
25	107.5	3.6	407	3	AAE50911	Aag50911 Arabidops

## ALIGNMENTS

### RESULT 1

AAE24846

ID AAE24846 standard; protein; 567 AA.

XX AC AAE24846;

XX XX

DT 30-OCT-2002 (first entry)

XX XX

DE KSHV latency-associated nuclear antigen 2 (LANA2).

XX XX

KW Kaposi's sarcoma-associated herpesvirus; KSHV; Castlemann's disease;  
KW latency-associated nuclear antigen 2; LANA2; apoptosis; tonsil tissue;  
KW primary effusion lymphoma; lymph node; spleen; skin lesion; virucide;  
KW cytostatic; antibody-based therapy; antisense therapy.

OS Human herpesvirus 8.

XX XX

FN WC200245744-A1.

XX XX

PD 13-JUN-2002.

XX XX

PF 07-DEC-2001; 2001WO-US047217.

XX XX

PR 08-DEC-2000; 2000US-00733728.

XX XX

PA (UYCO ) UNIV COLUMBIA NEW YORK.

XX XX

PI Chang Y, Moore PS;

XX XX

XX WPI; 2002-519642/55.

DR N-FSDB; AAD40494.

XX XX

PT Novel nucleic acid encoding Kaposi's sarcoma-associated herpesvirus

XX XX

PT latency-associated nuclear antigen 2 polypeptide for inhibiting p53

XX XX

PS mediated apoptosis of a cell.

XX XX

PS Claim 47; Fig 13; 11pp; English.

XX XX

CC The invention relates to an isolated nucleic acid which encodes a

XX XX

CC Kaposi's sarcoma-associated herpesvirus (KSHV) latency-associated nuclear

XX XX

CC antigen 2 (LANA2) polypeptide or its fragment. The invention is useful

XX XX

CC for inhibiting p53 mediated apoptosis of a cell, for immortalising a

XX XX

CC cell, preferably haematopoietic tissue cell or B cell and for producing

XX XX

CC an antibody. The invention is useful for treating Kaposi's sarcoma-

XX XX

CC associated herpesvirus infection in a subject. The invention is also is

XX XX

CC useful for determining whether a subject, including mouse, rat, dog,

CC guinea pig, ferret, rabbit, primate, or human being is afflicted with a

CC disease associated with Kaposi's sarcoma-associated herpesvirus infection

Aag50896 Arabidops  
Aae13003 Corn eIF-  
Aab61242 Murine br  
Abo32678 Secreted  
Adb90787 Mouse bre  
Aar93747 P. vivax  
Aaw97039 A. secrete  
Aag66528 Plasmodiu  
Abp66122 Bifidobac  
Aab59318 Drosophil  
Aae16790 Human tra  
Aay15103 Modified  
Adc06828 Human pro  
Aau75883 Human adh  
Adc06827 Human pro  
Aam39445 Human pol  
Aag64532 Human zin  
Abp41810 Human ova  
Aau15901 Human nov  
Abu54970 Human nov

26 107.5 3.6 433 3 AAG50896  
27 107 3.6 417 4 AAE13003  
28 107 3.6 883 4 AAB61242  
29 107 3.6 883 6 ABO32678  
30 107 3.6 883 7 ADB90787  
31 107 3.6 1018 2 AAR93747  
32 107 3.6 1018 2 AAW97039  
33 107 3.6 1018 4 AAG66528  
34 107 3.6 1033 5 ABP66122  
35 107 3.6 1061 4 AAB59318  
36 106.5 3.6 686 5 AAE16790  
37 104.5 3.5 503 2 AAY15103  
38 104.5 3.5 1973 7 ADC06828  
39 104.5 3.5 2047 5 AAU75883  
40 104.5 3.5 3863 7 ADC06827  
41 104 3.5 166 4 AAM39445  
42 104 3.5 170 4 AAG64532  
43 104 3.5 175 5 ABP41810  
44 104 3.5 195 4 AAU15901  
45 104 3.5 195 6 ABU54970

CC of a B cell, such as Castleman's disease and primary effusion lymphoma or  
CC whether a subject is infected with KSHV, by contacting tonsil tissue,  
CC lymph nodes, spleen, skin lesions, blood, serum, plasma cerebrospinal  
CC fluid, lymphocytes, urine, transudates, exudates, bone marrow cells, and  
CC detecting any antibody which is bound to LANA2, by ELISA, IFA or Western  
CC blotting. The invention is also useful for treating B-cell associated  
CC diseases such as Castleman's disease, and primary effusion lymphoma. The  
CC invention is useful in antibody-based therapy and antisense therapy. The  
CC present sequence is Kaposi's sarcoma-associated herpesvirus (KSHV)  
CC latency-associated nuclear antigen 2 (LANA2)  
XX  
SQ Sequence 567 AA;

Query Match 100.0%; Score 2993; DB 5; Length 567;  
Best Local Similarity 100.0%; Pred. No. 3.2e-295;  
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAGRLTWISEFIVGALDSKYPVWLDKSTGTFLAPARNVDVPLDSLOFIDFKREC 60  
Db 1 MAGRLTWISEFIVGALDSKYPVWLDKSTGTFLAPARNVDVPLDSLOFIDFKREC 60  
QY 61 LSKGLHPRDLGSPITAFKICTTSRRLRLPGEEVVGGINCRWRLLCAEVKECWVC 120  
Db 61 LSKGLHPRDLGSPITAFKICTTSRRLRLPGEEVVGGINCRWRLLCAEVKECWVC 120  
QY 121 VHARTLHSGSSILWEILYOHVSLEKRRRPRRPFVGENSDSEEDHAFCDVPVQTGA 180  
Db 121 VHARTLHSGSSILWEILYOHVSLEKRRRPRRPFVGENSDSEEDHAFCDVPVQTGA 180  
QY 181 ESDSGDEGPFSTRHSASGVQVDDANADSPGSDGEPSTRHSDSQPPPADETTVHTDNVE 240  
Db 181 ESDSGDEGPFSTRHSASGVQVDDANADSPGSDGEPSTRHSDSQPPPADETTVHTDNVE 240  
QY 241 DDLTLADKESACALMTHVGOEMDMLEAMCDEDLFDLLGIPEDVIATSPGSDTDSGVV 300  
Db 241 DDLTLADKESACALMTHVGOEMDMLEAMCDEDLFDLLGIPEDVIATSPGSDTDSGVV 300  
QY 301 TEGSIAASAVGAGVEDVYLAGALEAQNAGVEYVLEISDEVEDDAGLPPASRRRPVWGEF 360  
Db 301 TEGSIAASAVGAGVEDVYLAGALEAQNAGVEYVLEISDEVEDDAGLPPASRRRPVWGEF 360  
QY 361 LWDGDPRRHERPTTRIRKRSAYRVARPPVMTDLGLVEYVYFGRPAMSLVERKV 420  
Db 361 LWDGDPRRHERPTTRIRKRSAYRVARPPVMTDLGLVEYVYFGRPAMSLVERKV 420  
QY 421 FILCSQNLADISHSLCHSRKGLRVLLPKPDNNTPGSDVNLAAVLRSFASGLVIVSLR 480  
Db 421 FILCSQNLADISHSLCHSRKGLRVLLPKPDNNTPGSDVNLAAVLRSFASGLVIVSLR 480  
QY 481 SGIVYKQLCKSTVLYHGNNPPKPGVICGLSSRAVLDFVFNVAQYRIQGHIEHKKTTVF 540  
Db 481 SGIVYKQLCKSTVLYHGNNPPKPGVICGLSSRAVLDFVFNVAQYRIQGHIEHKKTTVF 540  
QY 541 GDPTSAEQFDMVPLVILKRLRSVTCDD 567  
Db 541 GDPTSAEQFDMVPLVILKRLRSVTCDD 567

RESULT 2  
ABB60323  
ID ABB60323 standard; protein; 777 AA.  
AC ABB60323;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 7761.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
OS Drosophila melanogaster.  
XX

PN WO200171042-A2.  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
PA  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR N-ESDB; ABL04426.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 7761; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB372072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 777 AA;

Query Match 4.2%; Score 125; DB 4; Length 777;  
Best Local Similarity 23.5%; Pred. No. 0.0078;  
Matches 80; Conservative 35; Mismatches 94; Indels 132; Gaps 20;  
QY 93 GEYEVVVGINCRRWLLCAEVK-----RCWCVHARTLHSGSSILWEILYQHSVR 143  
Db 102 GAELYKQEAAYR-RHCKTKCGEELPREESRPMCKCCY---TRFSSASN----- 146  
QY 144 LEKHRR-----RPRRPFVG-----ENSDSSEEDHAFCDVPVOTGAE 181  
Db 147 LSKHRSRPTCCQPEVDSPGSDGCKKCAFRKKNRDSDDDE-----TTSEE 196  
QY 182 SEDSGDEGPFSTRHSASGVQVDDANADSPGSDGEGPSTRHSDSQ-----PPPADE 231  
Db 197 SEDSDDDIPLASRLKTKLKQ-ESQNSD---SGDECFDFEPNNSDDADASGFLQPPPA-- 250  
QY 232 TVHTDNVEDDLTLDKESACALMTHVGOEMDMLEAMCDEDLFDLL-----GI-PEDVI 285  
Db 251 -MVKVEAFDEE---DFEYQDASMYVKTSTDTFSNE--KDKLLDVLNKGDLKPFESL 303  
QY 286 ATSPQGDITDASGVVTEGTSIAASAVAGVEDVYLAGALEAQNAGVEYVLEISDEVEDDGA 345  
Db 304 KVEQ-----GAGILDE--IAAVPLVEVAEDVLELRGHQMEK-- 338  
QY 346 GLPPASRRRPVGEFLWDDGPREHERPTTRIRI-----RHRK 381  
Db 339 --PP-----GPRKGRPPKEKIPVVRKRYK 362

RESULT 3  
AAB53186  
ID AAB53186 standard; protein; 364 AA.  
XX  
AC AAB53186;  
XX  
DT 28-FEB-2001 (first entry)  
XX  
DE Macaca mulatta rhadinovirus 17577 RRV R13 protein SEQ ID NO:129.

XX	Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus; genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6; IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine; cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma; lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy; splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.
XX	Macaca mulatta rhadinovirus 17577.
XX	WO200028040-A2.
XX	18-MAY-2000..
XX	05-NOV-1999; 99WO-US026260.
XX	06-NOV-1998; 98US-0107507P.
XX	20-NOV-1998; 98US-0109409P.
XX	(UYOR-) UNIV OREGON HEALTH SCI.
XX	Hong SW, Axthelm MK, Searles RP;
XX	WPI; 2000-376552/32.
XX	New rhesus rhadino virus for producing non-human primate model useful for testing potential treatments and efficacy of the candidate vaccine for conditions associated with RRV infection.
XX	Claim 5; Page 135; 141pp; English.
XX	The present invention describes a novel rhesus macaque rhadinovirus called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the RRV genome sequence, and AAB53123 to AAB53204 represent the proteins encoded by the genome sequence. The present invention also specifically claims the individual open reading frame (ORF) nucleotide sequences from the genome which encode the individual proteins, but these sequences are not given. A non-human animal infected with RRV can be used for testing the efficacy of drug in the treatment of condition associated with infection with RRV such as Kaposi's sarcoma, lymphoproliferative disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly, hypergammaglobulinemia or autoimmune haemolytic anaemia, by administering the drug to a immuno-compromised non-human primate preferably Rhesus macaque monkey obtained by as a result of infection by Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-human primate model for testing potential treatments for conditions associated with RRV infection. It is also useful for testing the efficacy of the candidate vaccine against RRV infection or conditions associated with its infection by administering the vaccine to the subject capable of infection with RRV, inoculating the subject with RRV and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213 represent sequence used in the exemplification of the present invention
XX	Sequence 364 AA;
XX	Query Match 4.1%; Score 122; DB 3; Length 364;
XX	Best Local Similarity 22.5%; Pred. No. 0.0047;
XX	Matches 82; Conservative 43; Mismatches 127; Indels 112; Gaps 19;
QY	246 LDKSACALMVHVGQEMDMLRWACDEBDFLLGLIPEDVIATSPQGGDTDASGVVTEGSI 305
DB	57 LDRDLECGHQHVNSCEKRLLEVRNAGEF---QDDARATTRG----- 99
QY	306 AASAVGAGVEDVYLAGALEAQNVAQGEVYVLEISDEEV-----DDGAGLPPASRR 353
DB	100 -----GERFFYLPAVDPLCYA--CILDOSHSTVLNLEAACVHGLEPFGTLPFP--- 146
QY	354 RVVGEFLWDCQPRHERPTTETRHKRLSAYYVAR-----RPVMTD--RLGVEV 404
DB	147 -PAPAE---ADG-----AARSVVARAARLATVAPPHDQDTFFWRLIRV 187
QY	405 FYFGRPAMSLVE-----RRVFILCSQNPLADISHSKRLGVLVLPKPD-DNNTGP 457

Db	188	YFYG----	SLVAHTSQDRRGRVLRHKQDKPK--GHCEFYG-TAYKXWLPKQLDGLPLP	244
Qy	458	GVNIIAAVLRSFASGLVIVSLRSGIVVXNLSCKSTVLYHGNNP-----	PKKF	504
Db	241	EQRETVCEIINGCEBVGFLHGNELGMVYDNRTRHTYRCAGNDAEGNHAQRAVSSVKSOI	300	
Qy	505	GVICGLSRRAVLVDENVNAQYRIQHE-HIKTTVTFIGDPTSAEQMDMPLVIKLRVSRV	563	
Db	301	FYVMGLLR-----LARSVPVGDTPFNSNAVTLYLGGREGSSKR-POVEVTIL-----	V	346
Qy	564	TCDD	567	
Db	347	ICQD	350	
RESULT 4				
AAB40574				
ID	AAB40574 standard; protein; 2518 AA.			
XX	AC	AAB40574;		
XX	DT	08-FEB-2001 (first entry)		
XX	DE	Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.		
XX	KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;		
XX	KW	vulnery; aniporiatic; antiparkinsonian; nootropic; neuroprotective;		
XX	KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;		
XX	KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
XX	KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
XX	KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
XX	KW	antanaemic; gene therapy; cancer; proliferative disorder; hypertension;		
XX	KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
XX	KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
XX	KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
XX	KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
XX	KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
XX	KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
XX	KW	thrombosis; contraceptive.		
OS	Homo sapiens.			
XX	PN	WO200058473-A2.		
XX	PD	05-OCT-2000.		
XX	PF	31-MAR-2000; 2000WO-US008621.		
XX	PR	31-MAR-1999; 99US-0127607P.		
PR	PR	02-APR-1999; 99US-0127636P.		
PR	PR	05-APR-1999; 99US-0127728P.		
PR	PR	30-MAR-2000; 2000US-00540763.		
PA	(CURA-) CURAGEN CORP.			
XX	PI	Shinkets RA, Leach M;		
XX	PP	WFI; 2000-602362/57.		
DR	DR	N-PSDB; AAC74783.		
XX	PT	Novel nucleic acids and peptides derived from open reading frame X,		
PT	PT	useful for treating e.g. cancers, proliferative disorders,		
XX	PT	neurodegenerative disorders and cardiovascular disease.		
XX	PS	Claim 11; Page 784-790; 5507pp; English.		
XX	CC	AAC74446 to AAC75606 encode the proteins given in AAB40237 to AAB43397,		
CC	CC	which represent the human ORFX open reading frames 1 to 3161. The OREF		
CC	CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;		
CC	CC	antiporiatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;		
CC	CC	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;		
CC	CC	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;		
CC	CC	dermatological; immunosuppressive; antiinflammatory; antibacterial;		

CC antiviral; antifungal; antitubercular; antithyroid; and antianemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
XX  
XX  
SQ Sequence 2518 AA;

Query Match 4.1%; Score 121.5; DB 3; Length 2518;  
Best Local Similarity 21.3%; Pred. No. 0.11;  
Matches 81; Conservative 48; Mismatches 132; Indels 119; Gaps 17;  
QY 105 RWRLLCAEV--KECWVCVHARTLHSGSSLEWILYQHSVLEKRRPRPFVGENSDS 162  
DB 633 RWSIAIEMVSGKTSQCKNFYFKQNLDEILQKHLMKERNARRKKKAPAAAS 692  
QY 163 SEEDHPAFCDVPTQ-----TGAESE-----DSGDE-----GPSTRHSAS 197  
DB 693 EEAAPP-----FVDEDEMEASGVSGNEEYVEAEALHAGSNEVPRGCGSPATVNSS 747  
QY 198 GYQ-----PVDDANAPSGDEGPTRHSDSQP-----PADETT----- 233  
DB 748 DTESIPSPHTEAAKTGONGPKPATLGADGPPGPPPTTPTTPTTPTTPTTPTTPTTPTT 807  
QY 234 -----VHTDNVEDDLTLDKESACALMYHVGQMDMLRAMCDLDFLL 278  
DB 808 TTPPAPPSAPPPVVPKKEE-----ETAAPVVEEGE-----QKPPAAELAVDT 856  
QY 279 GTPEDVIATS-----QPGGDTADGVTGSIASAVGAGVEDVYLALGALEAQNVA 329  
DB 857 GKAEFPVSECTEAEAGPAKGADEAATAEGALKAEKEGG-----SGRATTAKSS 910  
QY 330 G-----EYVLEISDEVEDDAG-----LPPASRRRPVVGDFLMDGPRRHERP-TTTRI 377  
DB 911 GAPPQSDSSATCSADEVEAEGDKNRLSPSPSLITGPD-----PRNAPSQKPLDL 964  
QY 378 RHRKLSAYRVARPPVMT 397  
DB 965 KOLKQRAA-----AIPPIQVT 980

RESULT 5  
AAB53182  
ID AAB53182 standard; protein; 253 AA.  
XX AAB53182;  
XX 28-FEB-2001 (first entry)  
DE Macaca mulatta rhadinovirus 17577 RRV R9 protein SEQ ID NO:121.  
XX  
XX Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;  
KW genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;  
KW IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;  
KW cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;  
KW lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;  
KW splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.  
XX  
XX Macaca mulatta rhadinovirus 17577.  
OS  
XX WO200028040-A2.  
XX  
XX PD 18-MAY-2000.  
XX  
XX 05-NOV-1999; 99WO-US026260.

XX 06-NOV-1998; 98US-0107507P.  
XX 20-NOV-1998; 98US-0109409P.  
XX (UYOR-) UNIV OREGON HEALTH SCI.  
XX  
XX Wong SW, Axthelm MK, Searles RP;  
XX WPI; 2000-376552/32.  
XX  
XX New thes rhadino virus for producing non-human primate model useful for  
XX testing potential treatments and efficacy of the candidate vaccine for  
XX conditions associated with RRV infection.  
XX  
XX Claim 5; Page 134; 141pp; English.  
XX

The present invention describes a novel rhesus macaque rhadinovirus  
CC called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the  
CC RRV genome sequence, and AAB53123 to AAB53204 represent the proteins  
CC encoded by the genome sequence. The present invention also specifically  
CC claims the individual open reading frame (ORF) nucleotide sequences from  
CC the genome which encode the individual proteins, but these sequences are  
CC not given. A non-human animal infected with RRV can be used for testing  
CC the efficacy of drug in the treatment of condition associated with  
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative  
CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,  
CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by  
CC administering the drug to an immuno-compromised non-human primate  
CC preferably Rhesus macaque monkey obtained by as a result of infection by  
CC Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-  
CC human primate model for testing potential treatments for conditions  
CC associated with RRV infection. It is also useful for testing the efficacy  
CC of the candidate vaccine against RRV infection or conditions associated  
CC with its infection by administering the vaccine to the subject capable of  
CC infection with RRV, inoculating the subject with RRV and observing the  
CC effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213  
CC represent sequence used in the exemplification of the present invention  
XX  
XX Sequence 253 AA;

Query Match 4.0%; Score 120; DB 3; Length 253;  
Best Local Similarity 22.5%; Pred. No. 0.0042;  
Matches 47; Conservative 32; Mismatches 90; Indels 40; Gaps 9;  
QY 383 RSAYRVAR-----PPVMTD--RLGVEVYFGRPAMSLVERKVFILCSQNPADIS 433  
DB 47 RSAYABARHLGVPEAPLPHQIVPFWRRLIQVYFVGLALDHTSQDRGVRLHPRVPHPG 106  
QY 434 HSLHSRKLRLVLLPKPDNDNTGPDVNLAAVLSRFSASGLVIVLSRGIYVNLCKSTV 493  
DB 107 HLCFYG-TGFTWFFSPDRKLTAEQITQIKTMLVAYNEGIVYHGNVTGVYVDNRNRET 165  
QY 494 LYHGNPPKPGVIGC--LSSRAVL-----DVFNVAQVR-----IQGHEHIKKTTF 538  
DB 166 YAAGND-----CNGDIQREVWFLSKQKTFNFMGFMKRLARSPGESHAPCNGATLY 217  
QY 539 IGGDPTSAEQFDMVPLVIVKLRSLRVTCDD 567  
DB 218 LSQQP-GAESPQVPTSV-----VVCQD 239

RESULT 6  
ABG22997  
ID ABG22997 standard; protein; 1316 AA.  
XX ABG22997;  
XX 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #22988.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.



XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Dmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS87184.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX PS Claim 20; SEQ ID NO 53356; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 1316 AA;  
 Query Match 3.9%; Score 117; DB 4; Length 1316;  
 Best Local Similarity 19.7%; Pred. No. 0.12;  
 Matches 122; Conservative 55; Mismatches 187; Indels 256; Gaps 29;  
 QY 24 LVKWLDRSTGTFIAPAARDNVIELDSIQFIDFK-----RECL-----SKGLHPRDLIG 72  
 DB 276 LYAW-DNRRETFKRSIGNDETDEKXKFLGFFKVKRNSKGCCLTTPNSPSMHSRLTL 334  
 QY 73 SPITAFGKICTTS-----RLRLRLPGEYEVVQGINCRWELLCAEYKECWCWCHART 125  
 DB 335 GPSLSLGSISGVSKEMKRRAPPFGSGPPQD-----KASE 373  
 QY 126 HLHSGSLWEILLQHSVRLBKRR-----PRRFVGENSSDSEHHPA-FCQVP 174  
 DB 374 KVSLSGS-----QIDLOKKRRAPAPPPPPPPSPPLIPNRTEDKEENRKRSTMVSLP 424  
 QY 175 VTGTGAESEDSGEGSTRHSASGVQVDDANADSPSGDGEPSRHSQPPADETTV 234  
 DB 425 L-----GSGSHCSPPGAPQVLSAEETVSGSCFASDTEIDSGVMSPSD-----IV 472  
 QY 235 HTDNVEDDLTLFLDK-----ESACALMYHVG----- 259  
 DB 473 SLDSQQDSMKYKDKWATQDCQDLQAGTFLDLPQKSPLWKEKNGSENHLRTEKAVTAS 532

QY 260 -QEMDLAR-----AMCDEDLFDLLGIPEDVIATSQ-----PG 291  
 DB 533 NDEEDLLIAGEPRKTLAELDELEEM-----EDSYETDTSSTSSIHGASNHCPQDAMIPH 588  
 QY 292 GDTASGVVTEGSIASAVGAGVEDVYLAGALEAQNAGVLEISDEVEDDGAGLPAS 351  
 DB 589 GDTDAIPVTFFIGEVSDDFDVSGLFS-----NRNNAGSF-----DSEGV--AS 629  
 QY 352 RR---RPVVGEFLMDGPRRHERPTTRIRHKLRKSAYRVARPVPMITDRLGVGVFYFG 408  
 DB 630 REDSLAPLQAE-----HSQP-----HEKARE-----EV----- 652  
 QY 409 RPAMSEVERKVFILCSQNPDLADISHSLHS-RKGLRVLVLPKPDNDNTGPGDVNLLAVL 467  
 DB 653 -PAL-----HPASHDVGKIRVAL-----SNISKQGNLMETAPRV 686  
 QY 468 RSFASGL-----VIVSLRSGIV-----VKNLCKSTVLVH-- 496  
 DB 687 TSFASNLTNDNLNAKVQDKVYGCADGERTQATERVNSQPVNEKDSNDKNAALAPTSHQR 746  
 QY 497 GNNPPKFGVICGLSSRAVL 516  
 DB 747 GQPGKSYRLKHGLTYKII 766  
 RESULT 7  
 AAM78838  
 ID AAM78838 standard; protein; 1007 AA.  
 XX AAM78838;  
 XX 06-NOV-2001 (first entry)  
 XX Human protein SEQ ID NO 1500.  
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX Homo sapiens.  
 XX WO200157190-A2.  
 XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US004098.  
 XX 03-FEB-2000; 2000US-00496914.  
 XX 27-APR-2000; 2000US-00560875.  
 XX 20-JUN-2000; 2000US-00598075.  
 XX 19-JUL-2000; 2000US-00620325.  
 XX 01-SEP-2000; 2000US-00654936.  
 XX 15-SEP-2000; 2000US-00663561.  
 XX 20-OCT-2000; 2000US-00693325.  
 XX 30-NOV-2000; 2000US-00728422.  
 XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y,  
 PI Ma Y, Zhao Qa, Wang J, Zhang J, Ren F, Chen R, Wang ZW,  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI; 2001-476283/51.  
 XX N-PSDB; AAK51971.  
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX Claim 20; Page 3790-3792; 6221pp; English.  
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM79323-AM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoietic regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM8020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
XX Sequence 1007 AA;

```

Query Match      3.8%; Score 115; DB 4; Length 1007;
Best Local Similarity 19.16%; Pred. No. 0.12;
Matches 112; Conservative 61; Mismatches 170; Indels 228; Gaps 25;

QY      146 KHRRRPR-----PFGENSDSSEHDPAFCVDPVQ-----TGAESD----- 184
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      185 ---SCDEGSPSTHSASGVQ-----VDDANADSPGSGDE-----GPSTRHSDQ- 225
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      64 VIIGQENPNNSASSEPRPEGHNPQVMDTEHNPDPDSSVPTDPTWGPERRGEESG 123
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      226 -----PPPADETTVHTDNVE----- 240
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      124 HELVAATGTAAGGGGLILASPKLGATPLPESTPAPPPPPPPPPPPGVGSHLNIPLIL 183
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      241 DLTLLDK-----ESAC---ALMVHYGQ-----EMDMLRAMWCDEDILFDLLG 279
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      184 EELRVLOQRQIQHWQMEQICRQVLLGSLGQTVGAPASPSLEFGTGTASSTKPLPLFS 243
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      280 IPEDV-----IATSPFGSDT-----DASGV----- 299
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      244 PIKPVQTSKTLASSSSSSSSSGAETPKQAFHLYHPLGSHQSPFSAGGVGRSHKTPAPS 303
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      300 -----VTEGSIASAASVGA--GVEDVVLGALAEQNAVCEYVLEISD 338
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      304 PALPGSTDOLIASHPLAPPSTTGLLAQCILGAARGLEATASPOLLKPNKSG---ELSY 359
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      339 EKVDDGAGLIPASRRRPVVGSEFLWDGP-----RRH--ERP-----TTR---RIR 378
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      360 GEWVGPLEKPGGRHKRCFCAKVFGSDSALQIHLRSHTGERPYKCNVCGNFRFTRGLMKVH 419
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      379 HKRLASAYRVARPPVMTIDRLGVEYVFQRP--AMSL-----VERKVFIL 423
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      420 FHRHREKYPHVQMPNPHVPPEHLVDVITSSGLPGMSVPPPEKAEAEAAATPGGVVERKPLVA 479
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      424 CSQNPLADI SHSCVLSHRKG-----LAVLLPK-PDDNNTGPG--DVNLLAAV 466
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      480 STTALSATESUTLSTAGTATAGCLPAFNKFLMKAVEPKNKADENTPPGSEGAISGV 539
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      467 LRSFASGL-----VIVLSRSGIYVKNLCST 492
      Db      : : : : : : : : : : : : : : : : : : : : : : : :
QY      540 AESSTATLMQLSKMLTSLPSWALLTNHFST 570
      Db      : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8	
ABU61812	
ID	ABU61812 standard; protein; 2507 AA.
XX	
AC	ABU61812;
XX	
DT	12-AUG-2003 (first entry)
XX	
DE	Human nuclear receptor corepressor SMRTe.
XX	
KW	Human; SMRTe; nuclear receptor corepressor; gene therapy; tissue typing;
KW	cancer.
XX	

OS	Homo sapiens.
XX	
XX	Key Location/Qualifiers
XX	Domain 166..429
FT	/label= SNC_domain
FT	166..206
FT	/label= Amphipathic helix
FT	/note= "Residues 172, 179, 186, 194 and 200 make up the hydrophobic heptad repeat"
FT	430..486
FT	/label= SANT_A_domain
FT	613..669
FT	/label= SANT_B_domain
XX	
XX	US2003027137-A1.
XX	
XX	06-FEB-2003.
XX	
XX	.27-MAR-2001; 2001US-00819104.
XX	
XX	29-MAR-2000; 2000US-0193138P.
PR	(CHEN/J) CHEN J D.
XX	
PA	Chen JD;
XX	
PI	WPI: 2003-466139/44.
XX	
DR	N-PSDB; AC62249, ACA62250.
DR	
XX	New SMRte proteins and nucleic acids, useful in gene therapy, predictive medicine, therapeutic or prophylactic treatment, chromosome mapping, tissue typing and in forensic biology.
PT	
PT	Claim 9; Fig 1; 90pp; English.
XX	
CC	The invention relates to an isolated SMRte nucleic acid molecule. The nucleic acids are useful in gene therapy, as hybridisation probes for identifying SMRte-encoding nucleic acid molecules and as primers for amplifying of SMRte nucleic acid molecules. The polypeptides are useful as immunogens to raise anti-SMRte antibodies. The SMRte molecules are useful as targets for discovering and developing modulating agents to regulate a variety of cellular processes, in screening assays, in predictive medicine, in therapeutic or prophylactic treatment, in chromosome mapping, tissue typing and in forensic identification of a biological sample. Modulators of SMRte are useful for treating or preventing a condition associated with aberrant SMRte protein or nucleic acid expression or activity, such as cancer. The present sequence represents the amino acid sequence of the human nuclear receptor corepressor SMRte
CC	
XX	
SQ	Sequence 2507 AA;
	Query Match 3.8%; Score 115; DB 6; Length 2507;
	Best Local Similarity 21.0%; Pred. No. 0.52;
	Matches 75; Conservative 53; Mismatches 139; Indels 90; Gaps 16;
QY	105 RNRLLCAEV--KECWCVHARTHLHGSSLWEILYQHVSVRLEKHRRPPRPVFGNSDS 162
Db	631 RNWSAIARMVGSKTVSQCNFYFNYYKRONQLDEILQQHKLCMEKERNARRKKKAPAAAS 690
QY	163 SEEDHPAFCDVPVTO-----TCAESDSDGCPSTRHSGASGVQ---PVDDANADS 209
Db	691 EEAAPP-----PVDEEHEAGSVTNE-EEMVEAEATVNNSDTESPSPFTEAKDT 744
QY	210 PGSGDGPGSTRHSDFQ-----FPADETTVHTDNV-EDDLTL----- 246
Db	745 GGNGPKFPATLGADGGPPGPPTPPPDIPATESATLAPTTPPPAFFPSSPPFVV 804
QY	247 -----DKESACALMVHVGOEMDMLEAMCDEDLFDLLGTFPEDVIANS-----QPQG 292
Db	805 PKEEKEETAAAPPVEEGE----QKPFAAEELAVDTGAESPVKGECTEEAEGPAKG 860
QY	293 DTDAAGVVTGEGIAASAVGAGVEDVYLAGEAQNVAG-----EYYLELSDEVDGAG- 346

DB 861 DAAEATAEALKAKEGG-----SGRATTAKSGAPOQSDSSATCSGADEVDEAGG 914  
 QY 347 -----LPPASRRPVVGGFLWDDGPRRHERP-TTTRIRHRLKRSAYVYRAPPVMIT 397  
 DB 915 DKRLLSRPSLLTFTGD-----PRNASQKPLDLKQKORAA-----AIPPIQVT 961

## RESULT 9

ABG25416  
 ID ABG25416 standard; protein, 1321 AA.

XX AC ABG25416;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #25407.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS89603.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX FS Claim 20; SEQ ID NO 55775; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1321 AA;

Query Match

Best Local Similarity 19.5%; Score 114.5; DB 4; Length 1321;

Matches 118; Conservative 61; Mismatches 213; Indels 213; Gaps 26;

QY 24 LVKWLDRSTGTFLAARNVDVPLDSLOPFIDFK-----RECL-----SKGLHPRDLIG 72  
 DB 279 LYAW-DNRRETRKSLGNDKDETDKXKFLGFKVKNRSNGCLTTPNSPMHSRLTL 337  
 QY 73 SPITAFKICTYS-----RRLRLPGEEYVVGGINCRRWRLLCAEYKECWCVHART 125  
 DB 338 GPSLSLSISGVSVKSEMKRRAPPVPGSGPPVQD-----KASE 376  
 QY 126 HLHSGSSLWEILYQHSVRLEKRR-----PRPFVGENSDSSSEEDHPA-FCDVP 174  
 DB 377 KVSLS-----QIDLQKKRRAPPPPPPPPLIENRTEDKEENKRSKTWVSLP 427  
 QY 175 VTQTGAESDSDGEGFSTRHSASGVQVDDANADSPGSGDEGFSRHSQPPPADETTV 234  
 DB 428 L---GSGSHCSPDGAPOVLSAEETVSGSCFASDSTTSDSGVMSSPSD-----IV 475  
 QY 235 HTDNVEDDLTLDK-----ESACALMVHV-----259  
 DB 476 SLDSQODSMKYDKWATDOEDCDDLAGTDLGPQKPLWEKNGSENSHLREKAVTAS 535  
 QY 260 -QEMDMLR-----AMCDEDLFDLLGIPEDVIATSQ-----PG 291  
 DB 536 NDEEDLLIAGEFRKTLAELDELEEM-----EDSYETDTSSTSSIHGASNHCPQDAMIH 591  
 QY 292 GDTDASGVVTGEGSIAASAVGVEDVYLAGALEAONVAGEYVLEISDEVDGAGLPPAS 351  
 DB 592 GDTDAIPVTFIGEVSDDPVDVSGLFS-----NRNNAGSF-----DSEGV--AS 632  
 QY 352 RR---RPVVGEFLWDDGPRRHERPTTRIRH---BKLSAYYRVARPPVMI-----396  
 DB 633 RRDSLAPLOAHSQPHKAREEVPALHPASHDVGGIRVALSNISKDGNLMETAPRTHN 692  
 QY 397 -----TDLRGEV--FYFG-----RPAMSLVERKVFILC-----SQ 426  
 DB 693 FALNLHTDNLNAKVXDKVYGCADGERTQATERVNSQPVNEKDSNDKNAALAPTSHQRQ 752  
 QY 427 NFLADISHSCLHSRKLRLVLPKPD-----DNTGTGVDVNLAAVLRSFASGLVIVS 478  
 DB 753 NP--GKSYRLKHGLTYKIIPPKSEMCVDRDVSLSUGAIDELGNLVSHPATGIRIIS 810  
 QY 479 LRSGI 483  
 DB 811 LSSSV 815

## RESULT 10

ADD43960  
 ID ADD43960 standard; protein, 1005 AA.

XX AC ADD43960;

XX DT 15-JAN-2004 (first entry)

XX DE Chlamydia trachomatis immunogenic protein, SEQ ID No 255.

XX KW immunogenic; infection; Chlamydia trachomatis; immune; vaccine;  
 gene therapy; antibacterial.

XX OS Chlamydia trachomatis.

XX FN WO2003049762-A2.

XX PD 19-JUN-2003.

XX PF 12-DEC-2002; 2002WO-IB005761.

XX PR 12-DEC-2001; 2001GB-00029732.

XX PR 06-AUG-2002; 2002GB-00018233.

XX PR 14-AUG-2002; 2002GB-00018924.

XX PA (CHIR-) CHIRON SPA.

PI	Grandi G, Ratti G;
XX	
DR	WPI; 2003-532882/50.
DR	N-PSDB; ADD43961.
XX	
XX	
PT	New immunogenic composition having a protein or encoding nucleic acid,
PT	useful for diagnosing, preventing and/or treating Chlamydia trachomatis
PT	infection.
XX	
XX	
PS	Claim 6; SEQ ID NO 255; 164pp; English.
XX	
CC	The invention relates to a novel immunogenic composition comprising a
CC	protein or nucleic acid, and an adjuvant, where the protein or nucleic
CC	acid comprises any of 131 fully defined amino acid or nucleotide
CC	sequences given in the specification, or has 50% or greater sequence
CC	identity to it, or their fragments. The protein and/or nucleic acid of
CC	the immunogenic composition is useful in the manufacture of a medicament
CC	for the treatment or prevention of infection due to Chlamydia
CC	trachomatis. The infection is treated or prevented by the medicament
CC	eliciting an immune response which is specific to a C. trachomatis
CC	elementary body, or for neutralising C. trachomatis elementary bodies,
CC	hence the immunogenic composition can be used in creating a vaccine. The
CC	immunogenic compositions can also be used for the diagnosis of C.
CC	trachomatis infection. The nucleic acids of the immunogenic compositions
CC	can be used to treat disorders by gene therapy. The immunogenic
CC	compositions have antibacterial activity. This sequence represents one of
CC	the 131 C. trachomatis proteins with immunogenic properties of the
CC	invention.
XX	
XX	
SQ	Sequence 1005 AA;

Query Match	3.8%;	Score 113.5;	DB 7;	Length 1005;	
Best Local Similarity	21.0%;	Pred. No. 0.17;			
Matches	58;	Conservative 41;	Mismatches 88;	Indels 89;	Gaps 13
Qy	176	TQTGAESDSG-----	DEGPGTRISASGVQ-PVDDANADSPGS-----	GDGEG-----	216
Db	491	TQAGPSSDDGIFSPNETPGAGAAASP	PPSPPIINNVNNGEITVNAIGTNTVNT	TNT	550
Qy	217	-PSTRHSUSQPPADETVHTDNVEDLTLDKESACA-----	LMYHVGQEMDM	266	
Db	551	TTPTQSTDASTDSIDIDINTNNQDINTTDXSDGAGGVNGDI	SETSSGDDSGSVS	610	
Qy	267	RAMCDED-----	LPDLGIPE-----	DVIATSPQGGTDA-----	SCWVTE 302
Db	611	SSESKNASVNGDGPAMKDLISAVRKHLDVVY	PCENGSGTEGPLPANQTLGDIVSDVENK	670	
Qy	303	GS-----	IAASAVGAGVEDVILAGALEACNVAGYVLEISDEEVDGGA	345	
Db	671	GSAQTKLSGNTAGGDDPTTTAAVNGABEITILS-----	DTDSGIGGDVSDTASSGDSGG	727	
Qy	346	GL-----	PPASRRPVPVGEFLWDGP-----	RRH 369	
Db	728	GVSFSSSSSNKNTAVG-----	NDGSPGLDILAAVRKH	759	

RESULT 11	
ABP96241	
ID	ABP96241 standard; protein; 1007 AA.
XX	
XX	ABP96241;
XX	
XX	
DT	19-MAY-2003 (first entry)
XX	
XX	Human nucleic-acid associated protein 24 SEQ ID NO:24.
DE	
XX	
KW	Human; nucleic-acid associated protein; NMAP; cardiant; cytostatic;
KW	neuroprotective; gene therapy; cardiovascular disorder; cancer;
KW	neurological disorder.
XX	
XX	
OS	Homo sapiens.
XX	
PN	WO2003016549-A2.

XX	27-FEB-2003.	
PD		
XX		
PF	14-AUG-2002; 2002W0-US025829.	
XX		
PR	17-AUG-2001; 2001US-031311P.	
FR	24-AUG-2001; 2001US-0314682P.	
PR	24-AUG-2001; 2001US-0314756P.	
PR	27-AUG-2001; 2001US-0315105P.	
PR	31-AUG-2001; 2001US-0316751P.	
PR	31-AUG-2001; 2001US-0316856P.	
PR	05-OCT-2001; 2001US-0328185P.	
XX		
PA	(INCY-) INCYTE GENOMICS INC.	
XX		
PI	Barroso I, Baughn MR, Becha SD, Blake JJ, Borowsky ML, Burford N;	
PI	Duggan BM, Elliott VS, Emerling BM, Forsythe IJ, Gietzen KJ;	
PI	Govard AE, Griffin JA, Hafalia AJA, Honchell CD, Ison CH, Khan FA;	
PI	Lai PG, Lee EA, Lee S, Lee SY, Li JX, Lu DAM, Lu Y, Lehr-Mason PM;	
PI	Nguyen DB, Ramkumar J, Sprague WW, Tang YT, Thangavelu K; Yue H;	
PI	Thornston M, Tran UK, Wallia NK, Warren BA, Xu Y, Yao MG, Yue H;	
PI	Yue H, Zebaradjian Y;	
XX		
DR	WPI; 2003-256709/25.	
DR	N-PSDB; ABZ79896.	
XX		
XX	New human nucleic acid-associated proteins polypeptide, useful for	
PT	preparing a composition for diagnosing or treating e.g., cardiovascular	
FT	or neurological disorders.	
XX		
PS	Claim 1; Page 240-242; 290pp; English.	
XX		
CC	ABZ79873 to ABZ79905 encode the human nucleic-acid associated proteins	
CC	designated NAAP-1 to NAAP-33 given in ABP96218 to ABP96250. The NAAP	
CC	sequences have cardiac, cytostatic and neuroprotective activities, and	
CC	can be used in gene therapy. The NAAP sequences can be used for preparing	
CC	a composition for diagnosing or treating a disease or condition	
CC	associated with decreased expression or overexpression of functional NAAP	
CC	e.g. cardiovascular or neurological disorders or cancer	
XX		
XX	Sequence 1007 AA;	
XX		
XX	Query Match 3.8%; Score 113; DB 6; Length 1007;	
XX	Best Local Similarity 19.4%; Pred. No. 0.2;	
XX	Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;	
QY	146 KERRRR-----PFVGENSDSSSEDPAFCDVPVQ-----TGASED-----184	
Db	4 RKQKPKQLISDCGSPASENGDASEDPHQVCAKCAQPTDTEFLAHQNACTPPVPM 63	
QY	185 ---SGDEGPFSTRSASGVQ-----VDDNADSPSGDE-----GPSTRHSDSQ- 225	
Db	64 VIIGQENPNNSASSEPREGEHNPNQVMDTSHNPDSGSSVPTDPTWGPERRGESSG 123	
QY	226 -----PPADFTTVTDNVE-----240	
Db	124 HFLVAATGTAAGGGGLLIASPGLKATLPPESTPAPPVPPPPPPPGVSGHLNIPLI 183	
QY	241 DDLTLDDK-----EGAC-----ALMYHVQ-----EMDMLRAMCDDLDLLG 279	
Db	184 EELRVLQQRQHQQWTEQICROVLLLSGLQTVGAPSPSELPCTGTASTTFLPLFS 243	
QY	280 IPEDV-----IATSPQGGDT-----DASGV-----299	
Db	244 PIKPVQTSKLTASSSSSSSSGAETPKOAFPHLYHPLGSHQHPFSAGVGRSHKTPAPS 303	
QY	300 -----VTEGSIAASAVGA--GVEDVVLACALBAONVAGEVVLISD 338	
Db	304 PALPGSTDQLIASPHLAFSTTGLLAQCLGAARGLEATASPGLLKPKNGSG-----ELSV 359	
QY	339 EYDDGAGLPPASRRPVPVGEFLWDDGP-----RRH--ERP-----TTR-----RIR 378	
Db	360 GYVWGPLEKPGGKRCFKCAKVGSGSALOILHRSHTGERPKYCNVCGNRFTFRGNLKVH 419	

QY 379 HRKLSAYRVARPPVMTIDELGVEVFGEPR-AMSL------VERKVFIL 423  
Db 420 FHRREKYPHVQMPHPVPEHLDVITSSGLFYGMSPPEKAEBAATPGGVERKPLVA 479  
QY 424 CSQNPADISHSCLHSRKG-----LRVLLPK-PDDNNTGFG-DVNLLAAV 466  
Db 480 STTALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNKADENTPPGSGSAISGV 539  
QY 467 LRSFAS-----GLVIVSLRSGIYVKNLCKST 492  
Db 540 AESTATRMQLSKLVTSLPSWALLTNHFKST 570

RESULT 12  
ID ADC06835 standard; protein; 1007 AA.  
AC ADC06835;  
DT 18-DEC-2003 (first entry)  
DE Human prostate cancer-related protein XM\_033473.  
KW cystostatic; prostate cancer; breast; gene therapy; transgenic; human.  
OS Homo sapiens.  
FN WO2003064599-A2.  
PD 07-AUG-2003.  
PP 24-JAN-2003; 2003WO-US001943.  
PR 25-JAN-2002; 2002US-00054935.  
PR 14-FEB-2002; 2002US-0356130P.  
PR 22-MAR-2002; 2002US-00102946.  
PR 08-APR-2002; 2002US-00117229.  
PR 14-MAY-2002; 2002US-00144198.  
PR 19-JUL-2002; 2002US-00197824.  
PA (ORIG-) ORIGENE TECHNOLOGIES INC.  
PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;  
DR WPI; 2003-679495/64.  
XX New isolated polynucleotide related to cancer genes, useful for  
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing  
PT or treating cancers, e.g. breast and prostate cancers.  
XX Disclosure; Fig 18; 128pp; English.  
XX The invention relates to a novel isolated polynucleotide comprising a  
CC differentially-regulated mammalian cancer gene. The polynucleotides of  
CC the invention demonstrate cytostatic activity and are differentially  
CC expressed in prostate cancer. The polynucleotide, polypeptides and  
CC methods of the invention may be useful for detecting, diagnosing,  
CC staging, monitoring, prognosticating, preventing or treating cancers,  
CC particularly breast and prostate cancers. Furthermore, the invention may  
CC be utilised during gene therapy procedures or in the production of  
CC transgenic animals. The current sequence is that of the prostate cancer-  
CC related protein of the invention.  
XX Sequence 1007 AA;  
SQ

Query Match 3.8%; Score 113; DB 7; Length 1007;  
Best Local Similarity 19.4%; Pred.No. 0.2;  
Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;  
QY 146 KHRPRPR-----PFVGENSDSEEDHFAFCDPVTO-----TGASED----- 184  
Db 4 RKORXPOQLISDCBPGSASEEDHPQVCAKCAQFTDPTPEFLAHQNACTPPVM 63

QY 185 ---SGDEGSTRHSASGVQP-----VDDANADSPSGDE-----GPSTRHSDSQ- 225  
Db 64 VIIGQENPNNSASSEPRPEGHNNPQVMDTEHSPNPDSSGSSVPTDPTWGPERRGEESG 123  
QY 226 -----PPADETTVHTDNVE----- 240  
Db 124 HFLVAATGTAAGGGGLILASPKLGATPLPESTAPPPPPPPPPGVSGLHAIPLIL 183  
QY 241 DDLTLLDK-----ESAC---ALMYHVGQ-----EMDMLMRAMCDEDLFDLLG 279  
Db 184 EELRVLOQRQHOMQMTQICRQVILLGLSGQTGAPASPSLPGTGTASTKPLPLIFS 243  
QY 280 IPEDV-----IATSPGGDT-----DASGV----- 299  
Db 244 PIKPVQTSKTLASSSSSSSSSSGASTPKQAFPHLXPLGSPHPSAGGVGRSHKTPAPS 303  
QY 300 -----VTEGSTAASAVGA--GVEDVYLAGALEAQNVAGEYVLEISD 338  
Db 304 PALPGSTDLIASPHLAFSTTGLLAAQCLGAARGLEATASPLGLKPKNGSG---ELSY 359  
QY 339 EEVDDGAGLPPASRRRPVVGELWDGP-----RSH---ERP-----TTR---R-R 378  
Db 360 GEVMGLEKPGGRHKRCFCAKVFGSDSALQIHLRSHTERPYKCNVCGNRFTRGNLKVH 419  
QY 379 HRKLSAYRVARPPVMTIDRLGVEVFGEPR-AMSL------VERKVFIL 423  
Db 420 FHRREKYPHVQMPHPVPEHLDVITSSGLFYGMSPPEKAEBAATPGGVERKPLVA 479  
QY 424 CSQNPADISHSCLHSRKG-----LRVLLPK-PDDNNTGFG-DVNLLAAV 466  
Db 480 STTALSATESLTLSTAGTATAPGLPAFNKFLMKAVEPKNKADENTPPGSGSAISGV 539  
QY 467 LRSFAS-----GLVIVSLRSGIYVKNLCKST 492  
Db 540 AESTATRMQLSKLVTSLPSWALLTNHFKST 570

RESULT 13  
ID AAM79822 standard; protein; 1019 AA.  
AC AAM79822;  
DT 06-NOV-2001 (first entry)  
DE Human protein SEQ ID NO 3468.  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX Homo sapiens.  
OS OS  
XX WO200157190-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 05-FEB-2001; 2001WO-US004098.  
PF  
XX 03-FEB-2000; 2000US-00496914.  
PR  
XX 27-APR-2000; 2000US-00560875.  
PR  
XX 20-JUN-2000; 2000US-00598075.  
PR  
XX 19-JUL-2000; 2000US-00620325.  
PR  
XX 01-SEP-2000; 2000US-00654936.  
PR  
XX 15-SEP-2000; 2000US-00663561.  
PR  
XX 20-OCT-2000; 2000US-00693325.  
PR  
XX 30-NOV-2000; 2000US-00728422.  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
PI

PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI	Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX	WPI; 2001-476283/51.
DR	N-PSDB; AAK52955.
DR	
XX	
PT	Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT	in diagnosis and gene therapy.
XX	
PS	Claim 20; Page 347-348; 6221pp; English.
XX	
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AMW78323-AMW80302) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC	(AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
CC	sequence listing were missing at the time of publication
XX	
XX	Sequence 1019 AA:
XX	Sequence 1019 AA:
XX	Sequence 1019 AA:

Query Match	3.8%;	Score 113;	DB 4;	Length 1019;	
Best Local Similarity	19.4%;	Pred. No. 0.2;			
Matches 111;	Conservative	61;	Mismatches 171;	Indels 228; Gaps 225	
QY	146	KHRRRPRR-----	PFGENSDSSEDDHAFACDVPVQ-----	TGAESD-----184	
Db	16	RKRXFPQLISDCEGFSASENGDAEEDHPQVCAKCAQFTDPTFLAHQNAACSTDPVW	75		
QY	185	---SGDEGSPTEHSAAGVOP-----	VDDANADSPSGDE-----	GPSTRUSDQ--225	
Db	76	VIIGQENPNSSASSEPRPEGHNNPQMDTEHSPDSSGVPTDPTGPPRRGESESG	135		
QY	226	-----	PPPADSETTVHTTONVE-----	240	
Db	136	HFLVAATGTAAGGGGLILASPKLGATPLPESTPAPPAPPPPPPPPGVSGHLNIPIL	195		
QY	241	DLTLTLDK-----	ESAC--ALMVHVGO-----	EMDMLRWAMCDEDFDLIG	279
Db	196	EBELVLOQRQHOMQMTQICROVLLIGSGIGQTVGAPASPELPGTGTASTKPLLP	255		
QY	280	IPEDV-----	IATSPQSGDT-----	DASGV-----299	
Db	256	PIKPVQTKTTLASSSSSSSSGAETPKQAFHLYHPLGSHQHPFSAGVGRSHKPTAPS	315		
QY	300	-----	VTEGSIASAAGCA--GVEDVYLACALQAQNVAGEYVLEISD	338	
Db	316	PALPGSTQDLIASPHLAPPSITGLLAAQCLGAARGLEATASPLGLAKPKNGSG--	371		
QY	339	ERVDDGAGLPASRRRPVUGFELMDDPG-----	RRH--ERP-----	TTR-----RIR	378
Db	372	GEWVGPLEKPGGRHKRCRCAKVFGSDSALQIHLRSHGTGSPYKCNVCGNRFTRGNLKVH	431		
QY	379	HKLRASATYVARPPVMTIDELGVEYVFYGRP--AMSL-	-----	VERKVFTIL	423
Db	432	PHRREKYPHYQMNPHVPPELHDVITSSGLPYGMSVPEPEKAEAEAAATPGGVGERKPLVA	491		
QY	424	CSQENPLADISHSLHSRKG-----	LRVLLPK--PDDNNKTGPG--DVNLLAAV	466	
Db	492	STTALSATESLTLTSTAGTATAGLPFAFKFVKMAVEPKNKADENTPFGSEGGAISGV	551		
QY	467	LRSFAS-----	GLVIVSRSGIYVKNLKST	492	
Db	552	ASBSATRMQLSKLVTSLPSWALLTNHFKST	582		

RESULT 14

ADE5522  
XX ADE5522 standard; protein; 466 AA.  
XX  
AC ADE5522;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
XX Rat Protein O55170, SEQ ID NO 1339.  
XX  
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
XX  
XX Rattus norvegicus.  
XX  
XX WO2003016475-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
XX  
XX 01-NOV-2001; 2001US-0346382P.  
XX  
XX 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO ) GEN HOSPITAL CORP.  
XX  
XX (FARB ) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
XX  
XX GENBANK; O55170.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a rat protein (shown in Table 2 of  
XX the specification) which is differentially expressed during pain. Note:  
XX the sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic form directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 466 AA;  
XX

```

Query Match      3.8%; Score 112.5; DB 7; Length 466;
Best Local Similarity 20.3%; Pred. No. 0.064;
Matches 83; Conservative 50; Mismatches 143; Indels 133; Gaps 21;

QY 119 WCWHAZ-----THHSG-----SSLNELLVQHSVRLEKRRRRPRFPVGNSD---S 162
      |||      |||      |||      |||      |||      |||      |||      |||

```

Db 114 WAQARRKLADQVPHLHNAELSKTLKWLRLNESD-----KRPTEBAERLRMQ 163  
 Qy 163 SEEDHPAFCDVPTQTG-----AESDSG---DEGPST-----RHASGVOP 201  
 Db 164 HKXDPDYKYQPRRRKNGKAQGEAECGGETDQGGAAIAQHYKSAHLDRHPESG-SP 222  
 Qy 202 VDDANADSPSGDEGSTRHSDSQPPADSTTV-----HTDNVED 241  
 Db 223 MSDGNPEHPSQSGHPT-----PPTPKTELQSGKADPKRDRSLGEGKPHIDFGNV 276  
 Qy 242 DLTLLDKESACALMYHVGQEMDMRAMCDEDLFDLLGIPEDVIATSPQ-----GUTDASGV 299  
 Db 277 DIGEISHE-----VMSNNMETFDVTELDQYLPNGHPGVSYSAAGY 318  
 Qy 300 VTGESIAASAVGAGVEDVYLALAGAEQNAVAGEYVLEISDEVDGAGLPPA-SRRPVPVG 358  
 Db 319 ---GLSALAVASG---HSAWISKPGVALPTV-----SPPAVDAKAQVKT 358  
 Qy 359 EFLWDDGPRRH-ERPTTRIRHRKLR-----SAYYRVARPPVMTDRLGVEYFYFGRPAM 412  
 Db 359 ETTGQGPFPHTDQPSQIAYTSLSLPHYGSAPPISRRPQFDYSDHQPSGP-YIGHAQ 417  
 Qy 413 SLEV-----ERKVFILCSQ-NPLADISHSLHSRKGRLVLLPKP 450  
 Db 418 ASGLYSAFSYMGPSQRPPLYTAISDPSGSPQSHSPHWEQPVVTTLSRP 466

## RESULT 15

AD55524  
 ID ADE55524 standard; protein; 466 AA.

XX AC ADE55524;

XX AC ADE55524;

XX DT 29-JAN-2004 (first entry)

XX DE Human Protein P56693, SEQ ID NO 1341.

XX KW Human; pain; neuronal tissue; Gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.

XX PN W02003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GCHO ) GEN HOSPITAL CORP.

XX PA (FARB ) BAYER AG.

XX PI Woolf C, D'urso D, Befort X, Costigan M;

XX PR WPI; 2003-268312/26.

XX DR GENBANK; P56693.

XX FT New composition comprising two or more isolated polypeptides, useful for

XX PT preparing a medicament for treating pain in an animal.

XX PS Claim 1; Page; 1017pp; English.

XX CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 466 AA;

Query Match 3.8%; Score 112.5; DB 7; Length 466;

Best Local Similarity 19.5%; Pred No. 0.064;

Matches 80; Conservative 46; Mismatches 148; Indels 137; Gaps 19;

Qy 119 WCVHAR-----TLHSG-----SSLWEILYQHSVRLEKRRRRPRPFVGENSD---S 162

Db 114 WAQARRKLADQVPHLHNAELSKTLKWLRLNESD-----KRPTEBAERLRMQ 163

Qy 163 SEEDHPAFCDVPTQTG-----TGAESDSGDEGPST-----RHASGVOP 201

Db 164 HKXDPDYKYQPRRRKNGKAQGEAECGGETDQGGAAIAQHYKSAHLDRHPESG-SP 222

Qy 202 VDDANADSPSGDEGSTRHSDSQPPADSTTV-----HTDNVED 241

Db 223 MSDGNPEHPSQSGHPT-----PPTPKTELQSGKADPKRDRSLGEGKPHIDFGNV 276

Qy 242 DLTLLDKESACALMYHVGQEMDMRAMCDEDLFDLLGIPEDVIATSPQGGDTASGVVT 301

Db 277 DIGEISHE-----VMSNNMETFDVTELDQYLPNGHPG-----H 309

Qy 302 EGSIAASAVGAGVEDVYLALAGAEQNAVAGEYVLEISDEVDGAGLPPAS-----RRRPV 356

Db 310 VSSYSAGYGLG-----SALAVASHSAWISK---PPGVALTVSPGVDAKAQV 356

Qy 357 VGEFLWDDGPRRH-ERPTTRIRHRKLR-----SAYYRVARPPVMTDRLGVEYFYFGRP 410

Db 357 KTETAGPQGPHTDQPSQIAYTSLSLPHYGSAPPISRRPQFDYSDHQPSGP-YIGHS 415

Qy 411 AMSLEV-----ERKVFILCSQ-NPLADISHSLHSRKGRLVLLPKP 450

Db 416 QASGLYSAFSYMGPSQRPPLYTAISDPSGSPQSHSPHWEQPVVTTLSRP 466

Search completed: June 9, 2004, 09:01:10

Job time : 65 secs





Db 361 LWDGPRRHERTTIRIRHKLRSAYRVARPPVMTDRLGVEVYFGRPAMSLVERKV 420  
QY 421 FILCSQNLADISHSCLSHRSGKRLVLLPKPDNDNTGPGVGNLLAAVLSFASGLVIVSLR 480  
Db 421 FILCSQNLADISHSCLSHRSGKRLVLLPKPDNDNTGPGVGNLLAAVLSFASGLVIVSLR 480  
QY 481 SGIVYKVLCKSVLVYHGNPPKPKFGVICGLSRVLDVFNVAQRIQGHIEHKKTTVFIG 540  
Db 481 SGIVYKVLCKSVLVYHGNPPKPKFGVICGLSRVLDVFNVAQRIQGHIEHKKTTVFIG 540  
QY 541 GDPTSAEQFDMVPLVILKLSRVTCDD 567  
Db 541 GDPTSAEQFDMVPLVILKLSRVTCDD 567

## RESULT 2

US-09-252-991A-19246  
; Sequence 19246, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19246  
; LENGTH: 479  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19246

Query Match 4.0%; Score 119.5; DB 4; Length 479;  
Best Local Similarity 21.4%; Pred. No. 0.0011;  
Matches 93; Conservative 49; Mismatches 163; Indels 129; Gaps 18;  
QY 172 DVFVTQTCAESDSDGDEGSTRHSAGVQVDDANADSPGSGDEGSTRHSDSP-----PP 228  
Db 3 EVFSTCGGSTRTRWRPGDSTR-STSPMSASSEANGPGRNVEKSTFNDDNGPFVLP 61  
QY 229 ADETTVHTDNVEDDLTLDDKESACALMVHVGQEMDML-----MRAMCDEDFDLILGIPEDV 284  
Db 62 LCRRAL-----VGIACRAAYPPAARDMAVASLSLHAAADDLFDLHGTGDDP 108  
QY 285 I-----ATSPQGGDT-----DASGVVTEGSIASAV 310  
Db 109 LDAGVHPADRLPHVAPAAQOLHAFVDHLAQOLGGEQLGFGDGLGQLLVVHVHAAV 168  
QY 311 GAGVEDVYLA--GALEAQNVAAGEVLEISDEEVD-----GAGLPPASR 352  
Db 169 GEDLHVLDLTHFGELEAG-----VLEVRDPAGLALLHVVDGRLQRRGPHRAGV 222  
QY 353 RPPVVGELWDGPRRHERPTTRIRHKLRSAYRVARPPVMTDRLGVEVYFGRPAM 412  
Db 223 GCALLGEHV-----HQVEEALGVAQVFLRHEHIVEEQLG-----GVL 261  
QY 413 SLEVERKVFILCSQNLADISHSCLSHRSGKRLVLLP-----KPDNDNTGP-----GDVNLAA 465  
Db 262 ALHAB-----LLQVAPALEALHAAHQEANGVLVRRIGLRDRDDHQVQQAQVGDHGLAV 316  
QY 466 VLRSFASGLVIVSLRSGIVYKVLCKSVLYHG-----NNPPKFGVICGLSSRAVLD 517  
Db 317 -----EQPVVALVHRGAHAQGVAGGRLGHGHDGLATDDPQAGLLIGAT---VFG 368  
QY 518 VFNVAQRIQ-CHE 530  
Db 369 DVPRAQRVQRDHE 382

## RESULT 3

US-09-489-039A-13221  
; Sequence 13221, Application US/09489039A  
; Patent No. 6610838  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breston et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 13221  
; LENGTH: 613  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13221

Query Match 3.6%; Score 108.5; DB 4; Length 613;  
Best Local Similarity 21.3%; Pred. No. 0.025;  
Matches 91; Conservative 36; Mismatches 149; Indels 151; Gaps 17;  
QY 148 RRRPREP-----FVGENSDSSEEDHPACDVPVTQTGAESDSDGEGPSTRHS 195  
Db 116 RHRPAQFGRILRSATLLAGRFAG-----YHPHA-----NAENCPPAAGSGGGDRS 159  
QY 196 ASGVQVDDANADSPGSGDEGSTRHSDSQ---PPPADETTVHTDNVEDDLTLDDKESAC 252  
Db 160 AA-----EPGAEQP---EPGQRYRHRPQRRHPADRPDRHRPVR-----198  
QY 253 ALMHVVGQEMDMLMRAMCDEDFDLILGIPEDVIATSPQCGDTDAGVVTGSGSIASAVGA 312  
Db 199 --QRHRRRPHLYR-----PDDAAYGALAGG--GRSSLVAAGHPARHPGPA 241  
QY 313 GV--EDVYLALAGALEAQNVAAGEVLEISDEEVDGAGLPASRRRPVVGFLWDGGRHE 370  
Db 242 AVCERDWPAGARRTAGIGGE---RLSRAGADLAGAPPAAGRWPI-----APSRRL 291  
QY 371 RPTTRIRHKLRSAYRVARPPVMTDRLGVEVYFGRP-----410  
Db 292 IASCLLLTAASLLIILGLAQGVPLTIDQVFSALFGDAPRVAMVWVNEWRLPRVLMALL 351  
QY 411 -AMSLVERKVFILCSQNLADISHSCLSHRSGKRLVLLPKPDNDNTGPGD-----459  
Db 352 IGNALGVSGAIFQSILTRNPLGS-----PDVMGFTGAWSGVLVAVLVFG 395  
QY 460 VNLLAAVLSFASGLV-----IVSLRSGIVYKVLCKSVLYHGNPPKFGVICGLSSR 513  
Db 396 QNLTAALAAAGGVLTSVLVWLLAWRNGIETFL-----IIIGIVR 438  
QY 514 AVLDFEN 520  
Db 439 AMLVAFN 445

## RESULT 4

US-09-347-833-8  
; Sequence 8, Application US/09347833  
; Patent No. 6294658  
; GENERAL INFORMATION:  
; APPLICANT: Famodu, Layo O.  
; APPLICANT: Odell, Joan T.  
; TITLE OF INVENTION: Factors Involved in Gene Expression  
; FILE REFERENCE: BB-1172  
; CURRENT APPLICATION NUMBER: US/09/347,833  
; CURRENT FILING DATE: 1999-07-02  
; EARLIER APPLICATION NUMBER: 60/092,415  
; EARLIER FILING DATE: July 10, 1998  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Microsoft Office 97

Query Match  
Best Local Similarity 21.3%; Pred. No. 0.019;  
Matches 90; Conservative 51; Mismatches 150; Indels 132; Gaps 20;

121 VHARTLHSGSSWEI-----PACDVPVTQTGAESDESGDEGSTRHSASGVQ-PVDDANADSPG 161  
122 VKAKEHDAFESWQIMSTWKLSSRVFMRDSDIDLRNKKQORRKBGPKKIEVHRD 81  
162 SSEDH-----LYQHSVRLEKRRRPRPFVGE-----NSD 161  
82 AAQERHAQSRGLGRPAVSVV---RRHPMDVYCPGPGFSASASSSSQSGIRGMPHPSRG 138  
212 SGDEGPTSTRSDPPPADETVHTDNVEDDILLKESACALMYHVGEMDMLRAMCD 271  
139 SODIRHERHQ-----FDRVTLPQRV-----KDEAITLGPOGGLARGMSIR--- 181  
272 EDLFDLIGIPEDVIATSQPGDITDASGVVTEGSIASAVGAGVEDVYLAGALEAQNVAGE 331  
182 -----GQP-----PVSNTIETPSVIDHRIIVSSNG-----YNSAAD 212  
332 YVLEISDEE-----VDGAGLPPASRRPVVGEFLWDDGPRRHERPTTIRHR----- 380  
213 WTSSGREGDSNRLPDRITSGRIPASSQSAVTS-----ORPASOEGRSRSKSYSE 261  
381 -KLRS-----AYV-RVAPPFMIIDRLGVVVFY-----FGRPAMSLSEVERK 419  
262 DELREKSVLITREYSNAKDEKVVLCIEELNANFVFLVSLWNDSFKERKMERELLAK 321  
420 VFI-LGQNPLADISHSLHSRK---GLRVLLPKPDN-NTGPDVNLAAVLRSFASGL 474  
322 LLVSLCSGR-----HNLKSKQSLDGLSNVLASLEDNLSDAPRATYXGLRLLARFVEES 375  
475 VIV 477  
376 ILL 378

## RESULT 5

US-08-072-610-2  
Sequence 2, Application US/08072610  
Patent No. 5532133

## GENERAL INFORMATION:

APPLICANT: Barnwell, John  
TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,  
Monoclonal Antibodies, and Diagnostic Assays  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby and Darby  
STREET: 805 Third Ave.  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10022-7513

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/072,610

FILING DATE: 19930602

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 5986/07686

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)527-7700  
TELEFAX: (212)753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium vivax  
IMMEDIATE SOURCE:  
CLONE: pMB3.3.1  
US-08-072-610-2

Query Match 3.6%; Score 107; DB 1; Length 1018;

Best Local Similarity 22.5%; Pred. No. 0.086;  
Matches 66; Conservative 33; Mismatches 92; Indels 102; Gaps 13;

QY 148 RRRPRPFVGENSDSS-----EEDHPAFCD-----VPVTTQGAESDS- 185  
Db 362 RRRNRNVEGETEAEAEVSEETPEGELEELEATPDDFDALDGTLEETBETAEGETV 421  
QY 186 -----GDEGSTRHSASGVQ-----PVDDANADSPGSGDEG----- 216  
Db 422 EGEEVTEGEVEGEAEAELEATPDDFQLEEPSGEGEGEGEGEALVAVP 481  
QY 217 ---PSTRHSOSOP-----PPADETVHTDNVEDDILLD----- 247  
Db 482 VVAEVEVVTPAQVKPMVAPTADS-TLFDVILNDLTYADITSFELPKQLKDPDAGE 540  
QY 248 -----KESACALMYHVG---QEMDMLRAMCDEDLFLLG---IPED---VIATSPGG 292  
Db 541 AVTPSKAPVQVAVGPAQGVTEELMQLQEDDFELEGTAEPAGEGELVLEGEGETE 600  
QY 293 DTDASGVVTEGSIASAVGAGVEDVYLAGALEAQNVAGEVYLE-ISDEVEDDG 344  
Db 601 BEPREGEPEGEVEEELATPDDF-----ELEPTGEVEETVEGEETAEG 648

## RESULT 6

US-08-719-822B-2

Sequence 2, Application US/08719822B

Patent No. 5874527

## GENERAL INFORMATION:

APPLICANT: Barnwell, John  
TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens  
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby and Darby  
STREET: 805 Third Ave.  
CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022-7513

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/719,822B

FILING DATE: 09/30/96

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 5986/17686US2

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)527-7700

TELEFAX: (212) 753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium vivax  
IMMEDIATE SOURCE:  
CLONE: pWE3.3.1  
US-08-719-822B-2

Query Match 3.6%; Score 107; DB 2; Length 1018;  
Best Local Similarity 22.5%; Pred. No. 0.086;

QY	148	RRRRPFPVGNSSDSS	-----	EDHAFCD	-----	VPVTOGARESDS	185
Db	362	RRRRNVGGEETEARAEGEVS	ETPGEEELATPDD	FALDGGTTLLEET	ETAGSETV	421	
QY	186	-----	QDEGFSTRHSASGVQ	-----	PVDANADSPGSGDEG	-----	216
Db	422	EGEETVEGEETVEGEAAE	EELEATPDDF	OLEPSPGEGEGEGEGE	GEALVAVP	481	
QY	217	-----	PSTRHSDSQP	-----	PPADETTHVNDVEDDLTLD	-----	247
Db	482	VVAEVEPVVTPAQPVKPM	APTAD	E-TLFPV	IDLNDLT	YADITSPEPLFKILKDP	540
QY	248	-----	KESACALMVHYG	-----	QENDMLNRACDEDLFDLLG	-----	I PED-----VIATSPQSG 292
Db	541	ATVPSKEAPVQVPVAVGP	QAVEPTEEL	MQLDQDDF	LELGTAEAP	EGELVLEGESEPT	600
QY	293	DTDAGVVTEGSI	AASAVGAGVEDPVY	LAGALEAQNVAGE	YVYLE	-ISDEVDDG	344
Db	601	PEPRGEPTGEVPEE	LEATPDDF	-----	ELEETGEVET	VEGEETAG	648

```

DB      601  EEPREGEPTGEVPEELEEPPDDF-----ELEETGEEVETVEGETAEG 648

RESULT 8
US-09-252-991A-21798
; Sequence 21798, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21798
; LENGTH: 1650
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21798

Query Match      3.5%; Score 105.5; DB 4; Length 1650;
Best Local Similarity 21.6%; Pred. No. 0.28;
Matches 92; Conservative 44; Mismatches 179; Indels 111; Gaps 18;

QY      166  DHPAPCDVPVVTQTGAESDSDGEGPSTRHSASGVQVPVDNADSPGSGDEGPFSTRHSDSQ 225

```

Query Match 3.5%; Score 105.5; DB 4; Length 1650;  
Best Local Similarity 21.6%; Pred. No. 0.28;  
Matches 92; Conservative 44; Mismatches 179; Indels 111; Gaps 18;  
QY 166 DHPAFCDVPVTQGAESDSDGEGPSTKHSAGVQVDDANADSPGSDGPGPSTHSDSQ 225  
||| . . . . .

592 DHPGQAVVQGLGDAQGAGDQAPAAVVOAGGEGEGAVAGDPPAGAVVHRAELAQOOR 651  
226 PPPADETVHTD-----NVEDDITLLDKESACALM--YHVGQMDMLMRACDEDFDILL 278  
552 AGRGQAAVVDQAAAEVEGDAGFADODASALVEAFQVGE-----QAL 697  
279 GIPEDVIATSPQGGDTDASGVVTVTSGSIAASAGVAGVEDVVLGAGALEAQNAGVGVLEISD 338  
698 GADPSLLAVVQPGGHQGDAGVAADAAVAVVQHAGADIHRTLGA---DHAGTAVVEAGA 753  
339 EVDVDDGAGLPAS-----RRPVVGE-----FLWDDGPRRHERPTTR--IRHR 380  
754 LORHAGIAEQPAALVWGLAGQORTGAGGEPATVVOAGARQAAAFADQRAALVVQHA 813  
381 KLSAYYRVARPPVPMITDRLGVEVYFGRPAMSLVERKVFILCSQNPADISHSLHSR 440  
814 AEAHAQAVLAEPTAV-----AVEQF-----AAVQQA-----VAPGQHPLG-LVQQALHGE 859  
441 KGLRV-----LLPKPDNNTGPGDVNLLAAV-----LRSFASGLVTVSL 479  
860 AQAAVADDLAAAVVQALLAGVHGDLRGAG--NLGAVVDLPRLSDAAAIRGQPGCLAVVD- 916  
480 RSGIYVKNLCKSTVLYHGNPNPKFGVTCGLSSRAVLDFVNVAGYRQCGHEHIKTTVFI 539  
917 RVGRDLQGLFAD-----QFATLLG-----QAAKRLQ-----VAL 946  
540 GGDPTS 545  
947 GGDTPS 952

RESULT 9

US-09-621-976-5191  
; Sequence 5191, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621.976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5191  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-5191

Query Match 3.5%; Score 104; DB 4; Length 164;  
Best Local Similarity 25.0%; Pred. No. 0.0083;  
Matches 38; Conservative 22; Mismatches 62; Indels 30; Gaps 7;

QY 100 QGINCRWRLLCAEVKECW---CVHARTLHSGSLWE-----ILYQHSV--- 142  
DB 19 QHVRQCK-----CLFEGH--WTYETGKKYLHRSPTAEKALKKEKENILLQOSIGET 72  
QY 143 RLEKRRRRPRPFV-----GENSDSEEDHPACDVPVQTGAESDSDGDEGSPTRHSAG 198  
DB 73 NVERKAKKRSKSVTSSSSSSSSSASDSSSESETSTSSSESDSDTDESSSSSSSSS 132  
QY 199 VQPVDDANADSPGSDGEPST---RHSDSQP 227  
DB 133 TTSSSSSSDSDSSSSSSSSSTSTSSSSDDEPP 164

RESULT 10

US-08-654-482-14  
; Sequence 14, Application US/08654482  
; Patent No. 6245562  
; GENERAL INFORMATION:

APPLICANT: Dalla-Favera, Riccardo  
TITLE OF INVENTION: IDENTIFICATION OF GENES ALTERED IN  
MULTIPLE MYELOMA  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/654,482  
APPLICATION NUMBER: US/08/654,482  
FILING DATE: 28-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 50995  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 451 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-654-482-14

Query Match 3.4%; Score 103; DB 3; Length 451;  
Best Local Similarity 22.4%; Pred. No. 0.059; 74; Indels 76; Gaps 12;  
Matches 52; Conservative 30; Mismatches 30; Indels 76; Gaps 12;  
QY 347 LPPASRRREVVGEEFLWDD-----GPRRHERPTTRIRHKLRSAYVR 388  
DB 171 MPFLDRS-----WRDVDPQHPPEIPYQCPMTFGRGHHWQGCACENGQVGTFFA 222  
QY 389 VARP-----PVMITD-----RLGVEVYFGRPAMSLVERKVFV--LCSQNP 428  
DB 223 CAPPSQAFGVPTEPSIRSAEALAFSDCLRHICLYY-----REILVKELTTSP 271  
QY 429 L-ADISHSCLHSRKG-LRVLLPKPDNNTGPGDVNLLAAVLRSFASGLVIVSLRSGIYVK 486  
DB 272 EGRISHGHGTYDASNLQDLVFPYEDNGHRKNIELLSHLR-----GVVLANPADGLYAK 327  
QY 487 NLCKSTVLVHG-----NNPPKFGVICGLSSRAVLDFVNVQY--RIQHEH 531  
DB 328 RLQSTIYWDGPLALCNDPRNK-----LERDQTKLFDTQQFLSELQAFAH 373

RESULT 11

US-09-252-991A-23689  
; Sequence 23689, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23689

LENGTH: 898  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23689

Query Match 3.4%; Score 102.5; DB 4; Length 898;  
Best Local Similarity 22.9%; Pred. No. 0.21;  
Matches 93; Conservative 43; Mismatches 152; Indels 119; Gaps 23;

QY 128 HSGSLWILYQHSVRLKRRRRPRPFVGENSDSSE--DHPAFCDVPVTTQGAESD- 184  
DB 352 HPG---RVLHHVAVGEGHLLRPQVQVQGRGEGFCGHVSLHGIPASCL-PESEDR 406  
QY 185 -SG--DEGSPSTRHSAGVQVVDNADSPGSG-----DEGSPSTRHSQPPADETVH 235  
DB 407 LSGKEKPAQRRGSA---FDVHGEAAAGFLVAFVHAGLAHG----- 449  
QY 236 TQNVEDDLTLDDKESACALMVHVGEMDLMRA-----MCDDELFDLIGIPEDVI-- 285  
DB 450 ----ADDLVERDEVLAQAQH-ARRVDGLHSHRVAFDAGHLDQPANVAGQEVVILHA 504  
QY 286 -----ATSQ-----PGSDTASGVVTEGTSAAASAVAGVEDVYLAGALBAQNVAG 330  
DB 505 DFGGVLDLHATAEHAERPGHRTGHPDLA-----LATDLGAGDRGVFELVE--DADRRGG 558  
QY 331 EYVLISDEEVDGAGLPPASRRPVVGEFLWDDGPRSE--RPTTRIRHKLRSAYR 388  
DB 559 E-----QEAHNAVVGDEARVVMON-----RRNDPRRAVGRKGH----- 595  
QY 389 VARPVMTIDRLGVZV-----FYFG-RPAMSLEYERKVFILCSQNPLADISHCLH 438  
DB 596 AAAAGVLLVHRQGEVDPVEHAESIAQAGLPLAQLAVERR-----RAAPDLQSGAGDALV 651  
QY 439 SRKGLRVLPK-PDDNNTGPGDVLNLAALV-----RSFASGLV 475  
DB 652 AAAGLDAILHLHPDQQAGAGFLGGAPGLLVGHOLADROQAFADAMV 698

RESULT 12  
US-08-431-080-28  
Sequence 28, Application US/08431080  
Patent No. 5698686  
GENERAL INFORMATION:  
APPLICANT: Gottschling, Daniel E.  
APPLICANT: Singer, Miriam S.  
TITLE OF INVENTION: Telomerase Compositions and Methods  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TEXAS  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/431,080  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SN 08/326,781  
FILING DATE: October 20, 1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:155/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000

TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1085 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-431-080-28

Query Match 3.4%; Score 102; DB 1; Length 1085;  
Best Local Similarity 18.9%; Pred. No. 0.33;  
Matches 60; Conservative 41; Mismatches 115; Indels 102; Gaps 11;

QY 145 EKRRRPRPPF-----VGENSDSSEEDHPAFCDVPVTTQGAESDSDGDEGSTRH 194  
DB 468 EQEERQKRLKTKTPKTRTTNNVDNEYIFNVFFQ-----SDDENSGHKSXKGRH 520  
QY 195 SASGVQVDDANADS---PGSGDEGPSTR-----HSDSQPP 227  
DB 521 K-SGKSHIHKXGKGNLIKSNDDLEPSTHVSNGKYDSSDDEYDNIILDVAMPDDE 579  
QY 228 PADETTVHTDNVEDDITLLDKESACALMVHVGQEMDLMRAMCDEBDFLLGIPEDVIAT 287  
DB 580 CSBSETSHDADTDEELRALDSDS-----LDIGTEL-----DDYED----- 615  
QY 288 SQPGGTDASGVV-----TEGSIARASAVGAGVEDVYLAGALEAQN 328  
DB 616 -----DDSSVNTNVIDIDDDPSFYHYSDSGS--SSLISSNDSKENSQSKCKHD 668  
QY 329 AGEVYLEISDEEVDGAGLPPASRRPVVGEFLWDDGPRRHERPTTRIRHKLRSAYR 388  
DB 669 LLBTVVVDESDTEDDNLPSPSRSKNIGS-----KAKEIVSSNVVGLAPPKLGTWE 721  
QY 389 VARPVMTIDRLGVZV 406  
DB 722 TDNKPFSIIDGLSTKSLY 739

RESULT 13  
US-08-938-534-28  
Sequence 28, Application US/08938534  
Patent No. 5916752  
GENERAL INFORMATION:  
APPLICANT: Gottschling, Daniel E.  
APPLICANT: Singer, Miriam S.  
TITLE OF INVENTION: Telomerase Compositions and Methods  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TEXAS  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,534  
FILING DATE: 28-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/431,080  
FILING DATE:  
APPLICATION NUMBER: SN 08/326,781  
FILING DATE: October 20, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:155/PAR

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1085 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-938-534-28

Query Match 3.4%; Score 102; DB 2; Length 1085;  
Best Local Similarity 18.9%; Pred. No. 0.33;  
Matches 60; Conservative 41; Mismatches 115; Indels 102; Gaps 11;  
QY 145 EKRRRRPRPF-----VGENSSSEEDHPAFCDVPVVTQTGAESDSGDEGPFSTRH 194  
DB 468 EQERKQRRLYKTKPSTRTTNSVDNDEYFNVFFQ-----SDDENSCHKSKKGRH 520  
QY 195 SASGVQPVDDANADS---PGSGDEGPSTR-----HSDSQPP 227  
DB 521 K-SGKSHIEHKNGKSNLKNDDLEPSTHSTVLSNGKYDSSDDEYDNLDDVAHMFSDDE 579  
QY 228 PADETTVHTDNVEDDLTLDKESACALMYHVGQEMDMLMRAMCDEDLFDLLGIPEDVIAT 287  
DB 580 CSESETSHDADTDELRALDSDS-----LDIGTEL-----DDYED----- 615  
QY 288 SQPGGDTASGVV-----TEGSIASAAGVAGVEDVYLAGALDAQNV 328  
DB 616 -----DDDDSSVTNVFIDLDLPDSFYHYDSDGS--SSLISNSDKENSCKDCKHD 668  
QY 329 AGEVYLEISDEEVDGAGLPASRRRPVVGFEFLWDDGPRRHERPTTTRIRHRKLSAYR 388  
DB 669 LLETVVVDDESTEDENLPPSSRSKNIGS-----KAKEIVSSNVVGLRPPKLGATWE 721  
QY 389 VARPPVMTIRLGVVEFY 406  
DB 722 TDNKPFSIIDGLSTKSLY 739

## RESULT 14

US-09-345-294-28  
Sequence 28, Application US/09345294  
Patent No. 6387619  
GENERAL INFORMATION:  
APPLICANT: Gottschling, Daniel E.  
Singer, Miriam S.  
TITLE OF INVENTION: Telomerase Compositions and Methods  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TEXAS  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/345,294  
FILING DATE: 30-Jun-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/431,080  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: ARCD:155/PAR

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1085 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 28:  
US-09-345-294-28

Query Match 3.4%; Score 102; DB 4; Length 1085;  
Best Local Similarity 18.9%; Pred. No. 0.33;  
Matches 60; Conservative 41; Mismatches 115; Indels 102; Gaps 11;  
QY 145 EKRRRRPRPF-----VGENSSSEEDHPAFCDVPVVTQTGAESDSGDEGPFSTRH 194  
DB 468 EQERKQRRLYKTKPSTRTTNSVDNDEYFNVFFQ-----SDDENSCHKSKKGRH 520  
QY 195 SASGVQPVDDANADS---PGSGDEGPSTR-----HSDSQPP 227  
DB 521 K-SGKSHIEHKNGKSNLKNDDLEPSTHSTVLSNGKYDSSDDEYDNLDDVAHMFSDDE 579  
QY 228 PADETTVHTDNVEDDLTLDKESACALMYHVGQEMDMLMRAMCDEDLFDLLGIPEDVIAT 287  
DB 580 CSESETSHDADTDELRALDSDS-----LDIGTEL-----DDYED----- 615  
QY 288 SQPGGDTASGVV-----TEGSIASAAGVAGVEDVYLAGALDAQNV 328  
DB 616 -----DDDDSSVTNVFIDLDLPDSFYHYDSDGS--SSLISNSDKENSCKDCKHD 668  
QY 329 AGEVYLEISDEEVDGAGLPASRRRPVVGFEFLWDDGPRRHERPTTTRIRHRKLSAYR 388  
DB 669 LLETVVVDDESTEDENLPPSSRSKNIGS-----KAKEIVSSNVVGLRPPKLGATWE 721  
QY 389 VARPPVMTIRLGVVEFY 406  
DB 722 TDNKPFSIIDGLSTKSLY 739

## RESULT 15

US-09-252-991A-32880  
Sequence 32880, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 32880  
LENGTH: 1224  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32880

Query Match 3.4%; Score 101.5; DB 4; Length 1224;  
Best Local Similarity 22.6%; Pred. No. 0.46;  
Matches 93; Conservative 30; Mismatches 166; Indels 123; Gaps 18;  
QY 143 RLEKRRRRPRPFVGENSSSEEDHPAFCDVPVVTQTGAESDSGDEG----- 189  
DB 58 RAQRDRSHRRPGRVRRRLRQAAEPA-----RTGVAGEESDPPGAPCGQRRRPAG 110  
QY 190 -PSTRHSAGVQPVDDANADSPGSGDEGFS-----TRHSDSQ 225





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 09:02:41 ; Search time 50 seconds  
(without alignments)  
3190.381 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRRITWSEFVIGALDSD.....QFDMVPLVIKLRVTCDD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/prodata/1/pubaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/prodata/1/pubaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/prodata/1/pubaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/prodata/1/pubaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/prodata/1/pubaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/prodata/1/pubaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/prodata/1/pubaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/prodata/1/pubaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/prodata/1/pubaa/US09\_PUBCOMB.pcp.\*  
10: /cgn2\_6/prodata/1/pubaa/US09\_PUBCOMB.pcp.\*  
11: /cgn2\_6/prodata/1/pubaa/US09\_PUBCOMB.pcp.\*  
12: /cgn2\_6/prodata/1/pubaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/prodata/1/pubaa/US10\_PUBCOMB.pcp.\*  
14: /cgn2\_6/prodata/1/pubaa/US10\_PUBCOMB.pcp.\*  
15: /cgn2\_6/prodata/1/pubaa/US10\_PUBCOMB.pcp.\*  
16: /cgn2\_6/prodata/1/pubaa/US10\_NEW\_PUB.pcp.\*  
17: /cgn2\_6/prodata/1/pubaa/US60\_NEW\_PUB.pcp.\*  
18: /cgn2\_6/prodata/1/pubaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.5	4.1	2517	12	US-10-087-192-654
2	119	4.0	691	12	US-10-425-114-45486
3	115	3.8	2507	10	US-09-819-104A-2
4	114	3.8	691	15	US-10-108-260A-4345
5	113	3.8	1007	16	US-10-197-824-38
6	112	3.7	1007	15	US-10-144-198-35
7	111	3.7	509	12	US-10-282-122A-61933
8	111	3.7	1216	15	US-10-028-248A-72
9	111	3.7	1216	15	US-10-107-782-72
10	108	3.6	810	12	US-10-425-114-40809
11	108	3.6	953	12	US-10-282-122A-64713
12	107	3.6	883	10	US-09-759-130B-342
13	107	3.6	883	14	US-10-189-123-72
14	107	3.6	883	14	US-10-188-495-72
15	107	3.6	1134	12	US-10-425-114-45409

16 106.5 3.6 686 15 US-10-297-022-27 Sequence 27, Appl  
17 106.5 3.6 686 15 US-10-384-919-17 Sequence 17, Appl  
18 106.5 3.6 691 12 US-10-072-012-104 Sequence 104, App  
19 105 3.5 402 12 US-10-425-114-54050 Sequence 54050, A  
20 105 3.5 656 12 US-10-425-114-63586 Sequence 63586, A  
21 104.5 3.5 1161 12 US-10-307-817-554 Sequence 554, App  
22 104.5 3.5 1973 16 US-10-197-824-34 Sequence 34, Appl  
23 104.5 3.5 2047 12 US-10-346-963-2 Sequence 2, Appl  
24 104.5 3.5 3863 16 US-10-197-824-7 Sequence 7, Appl  
25 104 3.5 175 15 US-10-264-049-2942 Sequence 2942, App  
26 104 3.5 195 9 US-09-764-864-854 Sequence 854, App  
27 104 3.5 323 12 US-10-412-699B-1510 Sequence 1510, App  
28 104 3.5 323 15 US-10-374-780A-1400 Sequence 1400, App  
29 104 3.5 363 12 US-10-425-114-41846 Sequence 41846, A  
30 104 3.5 641 14 US-10-128-714-3031 Sequence 3031, App  
31 104 3.5 717 14 US-10-128-714-8031 Sequence 8031, App  
32 104 3.5 852 15 US-10-369-493-13138 Sequence 13138, A  
33 103.5 3.5 322 15 US-10-374-780A-1401 Sequence 1401, App  
34 103 3.4 299 12 US-10-424-599-254370 Sequence 254370, A  
35 103 3.4 357 12 US-10-282-122A-62763 Sequence 62763, A  
36 103 3.4 357 12 US-10-282-122A-64499 Sequence 64499, A  
37 103 3.4 629 15 US-10-384-919-2 Sequence 2, Appl  
38 103 3.4 2766 12 US-09-964-956-62 Sequence 62, Appl  
39 102.5 3.4 542 15 US-10-120-801-14 Sequence 14, Appl  
40 102 3.4 603 12 US-10-425-114-60357 Sequence 60357, A  
41 101.5 3.4 301 12 US-10-425-114-46877 Sequence 46877, A  
42 101.5 3.4 1239 10 US-09-291-417-13 Sequence 13, Appl  
43 101 3.4 660 12 US-10-425-114-52062 Sequence 52062, A  
44 101 3.4 1000 14 US-10-128-714-3305 Sequence 3305, App  
45 100.5 3.4 411 12 US-10-425-114-65986 Sequence 65986, A

#### ALIGNMENTS

RESULT 1  
US-10-087-192-654  
; Sequence 654, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 654  
; LENGTH: 2517  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-087-192-654

Query Match 4.1%; Score 121.5; DB 12; Length 2517;  
Best Local Similarity 21.3%; Pred. No. 0.28;  
Matches 81; Conservative 48; Mismatches 132; Indels 119; Gaps 17;  
QY 105 RWELLCAEV--KECWVCHVARTLHSGSSWELLYOHSVRLKCHRRPRPFGVNSDS 162  
DB 632 RNWAIARMGSKTVSCKNFYFKYKQNLDELLOQHKLMKERNARKKKAPAAAS 691  
QY 163 SEEDHPAFCDVPVTQ-----TGAES-----DSGDB-----GPSTRHAS 197  
DB 692 EEAFAFP-----PVVEDSEMEASGVSGNEEVEAEALHASGNEVPGECGATVNSS 746  
QY 198 GVQ-----PVDANADSPSGDEGPSTRHSQPP-----PADETT--- 233

Db 747 DTESIPSPHTEAAKOTGONGKPPATLGADGPPGPTPPEDIPAPTEPTASEATGAP 806  
QY 234 -----VHTDNVEDDLTLTKSACALMVHVGQEMDMLRAMCDEDLFLL 278  
Db 807 TTPPAPPSPSAPPVVPKEEKE-----ETAAAPPVEEGEE-----QKPPAAEELAVDT 855  
QY 279 GTPEDVIATS-----QPGGDTASGVVTEGSIASAVGAGVEDVYLAGALEAQNV 329  
Db 856 GRAEFPVKSECTEEAEFGPAKDAEAAEATAEGALKAKEGEG-----SGRATTAKSS 909  
QY 330 G-----EYVLEISDEVEDDAG-----LPPASRRPVVGVFLWDDGPRRHERP-TTTRI 377  
Db 910 GAQSDSSATCSADEVEAGDKNRLLSRPSSLITPTGD-----PRANASQKPLDL 963  
QY 378 RHRKLSAYRVARPPVMIT 397  
Db 964 KOLKQRAA-----AIPPIQVT 979

RESULT 2  
US-10-425-114-45486  
; Sequence 45486, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 45486  
; LENGTH: 691  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700346396\_FLI.pap  
US-10-425-114-45486

Query Match 4.0%; Score 119; DB 12; Length 691;  
Best Local Similarity 19.6%; Pred. No. 0.065;  
Matches 71; Conservative 58; Mismatches 157; Indels 76; Gaps 15;

QY 72 GSPITAP-----GKICTTERRLR--RLPGEYEVVQGINCRWRLLCAEVKECW 119  
Db 138 GSSVEFPVNSFVQNLQKSFIDEKLRVWRVAEFSVPPDSRALQWKITCGSNK--- 194  
QY 120 CVHARTHLHSGSLWEILYQHSVRLEKHRRRPR--PFVGENSDSSSEEDHPAFCDVPVTQT 178  
Db 195 --HG---LPKSSLLKQEMERGDKVDRHAVHERKSKAMPEGYQKQENNVKPKDIIHVVD 249  
QY 179 GAESDSGDEGSPSTRHSASGVQVDDANADSPGSGDEGSPSTRHSDSQPPPADETTVHTDN 238  
Db 250 GI-----QQLGCKNRKNYS-----DKPSERRHNDNPLPLD---MKEKN 285  
QY 239 VEDDLTLTKSACALMVHVGQEMDMLRAMCDEDLFLLGIPEDVIATSPQGGDTASG 298  
Db 286 GCKM-----KKS-----GKQSDRRRELMDAEVLIDINGLKQDVTGKPKSGGPRSW 333  
QY 299 VVTEGSIASAVGAGVEDVYLAGALEAQNV-----AGEYVLEISDEVEDDAGLP--PA 350  
Db 334 --GHADLGLTKLTGIEKQIDISTLINGKAVNKPYPYKPYRAPMGKEKVAEDRQVPPEKAT 391  
QY 351 SRRRPVVG---EFLWDDGPRRHERPTTIRIRHKLRSAY-----YRVARPPVMITRLG 401  
Db 392 NNRPPYKPNFKHANQANGHOSGTEETHOKREPIYDPVSVRSRIPRPPAHADYAG 451  
QY 402 VE 403

Db 452 ME 453  
:|

RESULT 3  
US-09-819-104A-2  
; Sequence 2, Application US/09819104A  
; Publication No. US20030027137A1  
; GENERAL INFORMATION:  
; APPLICANT: Chen, J. Don  
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES  
; TITLE OF INVENTION: AND USES THEREFOR  
; FILE REFERENCE: UMG-030  
; CURRENT APPLICATION NUMBER: US/09/819,104A  
; CURRENT FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/193,138  
; PRIOR FILING DATE: 2000-03-29  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2507  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-819-104A-2

Query Match 3.8%; Score 115; DB 10; Length 2507;  
Best Local Similarity 21.0%; Pred. No. 1.2;  
Matches 75; Conservative 53; Mismatches 139; Indels 90; Gaps 16;

QY 105 RWRLLCAEV--KECWCVHARTHLHSGSLWEILYQHSVRLEKHRRRPRRPFVGENSDS 162  
Db 631 RWSAIARWVGSKTVSQCKQNFYNYKQNLDEILQOQHLKWKERNARKKKKAPAAAS 690  
QY 163 SEEDHPAFCDVPVTQ-----TCASESDSGDEGSPSTRHSASGVQ-----PVDDANADS 209  
Db 691 EBAAPF-----FVDEDEMEASGVGTGNE--EMVEEAETVNNSSDTSIPSPHTEAAKDT 744  
QY 210 PGSGDEGSPSTRHSDSQP-----PPADETTVHTDNV-EDDLTL----- 246  
Db 745 GONGKPPATLGADGPPGPTPPEDIPAPTEPTASEATLAPPPAPPFPSPPPV 804  
QY 247 -----DKSACALMVHVGQEMDMLRAMCDEDLFLLGIPEDVIATS-----QPGG 292  
Db 805 PKEEKEETATAAPPVEEGEE---QKPPAAEELAVDTGKAEPVKSECTEEAEFGPAK 860  
QY 293 DTDASGVVTEGSIASAVGAGVEDVYLAGALEAQNVAG-----EYVLEISDEVEDDAG- 346  
Db 861 DAAAEATAERALKAKKEGG-----SGRATTAKSSGAPQSDSSATCSADEVDEAEGG 914  
QY 347 -----LPPASRRPVVGVFLWDDGPRRHERP-TTIRIRHKLRSAYRVARPPVMIT 397  
Db 915 DKNRLLSRPSSLITPTGD-----PRANASQKPLDLKQLKQRAA-----AIPPIQVT 961

RESULT 4  
US-10-108-260A-4345  
; Sequence 4345, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4345  
; LENGTH: 691  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4345

Query Match 3.8%; Score 114; DB 15; Length 691;



Db 540 AESSTATRMQLSKLVTSPLSPWALLTNHFKST 570

RESULT 7

US-10-282-122A-61933  
; Sequence 61933, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 61933  
; LENGTH: 509  
; TYPE: PRT  
; ORGANISM: Mycobacterium avium  
US-10-282-122A-61933

Query Match 3.7%; Score 111; DB 12; Length 509;  
Best Local Similarity 24.5%; Pred. No. 0.24;  
Matches 54; Conservative 31; Mismatches 65; Indels 70; Gaps 12;  
QY 172 DVPVTQTGAEEDESGDEGSPRHSASGVQPVYD-----DANADSPGSGDEGSPRHS- 222  
Db 11 DGPVTRATKSPSPAKPAK-AANGSAPAKRATKATRSASKEACAAEPAKTRSSA 69  
QY 223 ---DSQPP-----PAETTVHTDNVEDDL-TLLDKESACALMYHVGQEMDLMR 267  
Db 70 KGADAKAPSGRGTRAAKGPADPALDTGAVEDLDTEPDLEG-----EPGEDLDI--- 120  
QY 268 AMCEDDFDLIGIPEDVIATIS---OPGDDTDASGVVTEGSTAASVACGVEDVYLAGALE 324  
Db 121 ---DTDL-NLDDLEEDVAADADIEPG---DAEGEDEEAAPKAAGATAAD----- 165  
QY 325 AQNVAGETVLEISDEEVDGAGLPPASRRPVPVGEFLWDD 364  
Db 166 -----EDEIAEPSEKDKAS-----GDFWDE 187

RESULT 8

US-10-028-248A-72  
; Sequence 72, Application US/10028248A  
; Publication No. US20030235882A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Vernet, Corine  
; APPLICANT: Casman, Stacie  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Miller, Charles  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Li, Li  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Colman, Steven  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Si, Jingsheng  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Stone, David  
; APPLICANT: Sciore, Paul  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Rothenberg, Mark  
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods  
; TITLE OF INVENTION: Thereof  
; FILE REFERENCE: 21402-222  
; CURRENT APPLICATION NUMBER: US/10/028,248A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256619  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/262959  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/272408  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/285189  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/308039  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: 60/311266  
; PRIOR FILING DATE: 2001-08-09  
; NUMBER OF SEQ ID NOS: 211  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 72  
; LENGTH: 1216  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-10-028-248A-72

Query Match 3.7%; Score 111; DB 15; Length 1216;  
Best Local Similarity 20.4%; Pred. No. 0.91;  
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;  
QY 30 RTGTGFLPAAANDVPLDLSLOFFIDFKRCLSKGLHPRDLGLSPITAFGKICTSRRLR 89  
Db 708 RQGGALLTRPQVPVVP-----KECLSFMYH---LHGPOI---GTLCLAMRR-- 748  
QY 90 RLPGGEYEVW---QGINCRRWRLLCAEVKEQWCVCVHARTHLHSGSSLWEIL-----YQ 139  
Db 749 ---EGEDILLNSRSTGHNWH-----QAWVTLH---HQLPSTKYQLLFEGLRDCYH 796  
QY 140 HSVRLKHKRRPRRRFVGENSDSSEEDHPACDVPVVTGTGAESDSDGEGPSTR-HSASG 198  
Db 797 GTMGDDMAVRPGPCWAAKRCFSFSD---CGF-----SPGDWGLWTRQNNASG 842  
QY 199 VQP-----VDDANADSPGS---GDEGP-----STRHSDSQPPPADETTVHTDNVEDDL 243

Db 843 LGPWGPIWIDTTGTAQGHYVMDTSPNLLPKGHVASLTSEHPP----- 886  
QY 244 TLLDKESACALMYHVGQEMDMRAMCDEDLFLGIPEDVIATSPGGDTSAGVVT-- 301  
Db 887 --LSRPACLSFWHLSPHNPCTLRVFE-----STRQELSIHGFGFAWRLGSVNVQ 938  
QY 302 ---EGSIAASAGAGVEDVYLAGALEAQNAGVYVLEISDEEYDDGAGLPPAS----- 351  
Db 939 AEQAWKVVFEMASGVHSYNA-----LDDISLQDGFCAQPGSCDFESGL 983  
QY 352 -----RRPVVGEFLND-----DGPRHERPTTRIRHKLRSAYYRVARPPVMTDRLG 402  
Db 984 CGWSHLPWPGLGYSWDSGATPSRYPRPS-----VDHTVGT 1021  
QY 403 EVFYF-----GRPAMSLEVERKVFIILCSQNPDIADISHSCL-----HSRKG 442  
Db 1022 EAGHPAFFETSVLPGGQAALGSE-----PLPATAVSCLHFWYMGFFAIFYKG 1071  
QY 443 -LRVLL 447  
Db 1072 ELRVLL 1077

## RESULT 9

US-10-107-782-72  
; Sequence 72, Application US/10107782  
; Publication No. US20040018970A1  
; GENERAL INFORMATION:  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Casman, Steve  
; APPLICANT: Coleman, Steve  
; APPLICANT: Edinger, Shomit,  
; APPLICANT: Gangoli, Bha,  
; APPLICANT: Kekuda, Ramesh,  
; APPLICANT: Li, Li,  
; APPLICANT: Liu, Xiaohong,  
; APPLICANT: Malyankar, Uriel,  
; APPLICANT: Miller, Charles,  
; APPLICANT: Millet, Isabelle,  
; APPLICANT: Patturajan, Meera,  
; APPLICANT: Rothenberg, Mark,  
; APPLICANT: Sciore, Paul,  
; APPLICANT: Shenoy, Suresh,  
; APPLICANT: Shinkets, Richard,  
; APPLICANT: Si, Jingsheng,  
; APPLICANT: Smithson, Glennda,  
; APPLICANT: Spytek, Kimberly,  
; APPLICANT: Stone, David,  
; APPLICANT: Taupier, Raymond, jr.,  
; APPLICANT: Tchernev, Velizar,  
; APPLICANT: Vernet, Corine,  
; APPLICANT: Zerhusen, Brian  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF  
; FILE REFERENCE: 21402-222CIP  
; CURRENT APPLICATION NUMBER: US/10/107,782  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: 10/028,248  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/256,619  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/262,959  
; PRIOR FILING DATE: 2001-01-19  
; PRIOR APPLICATION NUMBER: 60/272,408  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/285,189  
; PRIOR FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/308,039  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: 60/311,266  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/279,344  
; PRIOR FILING DATE: 2001-03-28

; NUMBER OF SEQ ID NOS: 215  
; SOFTWARE: Curaseq1 version 0.1  
; SEQ ID NO 72  
; LENGTH: 1216  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-107-782-72

Query Match 3.7%; Score 111; DB 15; Length 1216;  
Best Local Similarity 20.4%; Pred. No. 0.91;  
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;

QY 30 RSTGTFLAPARNDIVPLDSLOFFIDFKRECLSGKLHPRDLGSPITAFGKICTTSRLR 89  
Db 708 RQGGALLTRPQVPVP-----KECLSWYH---LHGPOI---GTLCLAMRR-- 748  
QY 90 RLPGEYEW--QGINCRRWRLCAEVKECWCWCVHARTHLHSGSSSLWEIL-----YQ 139  
Db 749 --EGEEDTLLWSRSGTHGRWH-----QAWTLH---HQLPSTKYQLLFEGLRDCYH 796  
QY 140 HSVRLKHHRRPRPRPFVGENSDSSEEDHPAFCDVPVVTQTGAESDSDGDEGSTR-HSASG 198  
Db 797 GTYGLDDMAVRPGPCWAAKRCSEFSD---CGF-----SPGDWGLWTRQNNASG 842  
QY 199 VQP---VDDANADSPGS---GDEGP-----STRHSDSQPPPADETTVHTDNVEDDL 243  
Db 843 LQPGFWIDHTTGTAGHYMVVDTSPNLLPKGHVASLTSEHPP----- 886  
QY 244 TLLDKESACALMYHVGQEMDMRAMCDEDLFLGIPEDVIATSPGGDTSAGVVT-- 301  
Db 887 --LSRPACLSFWHLSPHNPCTLRVFE-----STRQELSIHGFGFAWRLGSVNVQ 938  
QY 302 ---EGSIAASAGAGVEDVYLAGALEAQNAGVYVLEISDEEYDDGAGLPPAS----- 351  
Db 939 AEQAWKVVFEMASGVHSYNA-----LDDISLQDGFCAQPGSCDFESGL 983  
QY 352 -----RRPVVGEFLND-----DGPRHERPTTRIRHKLRSAYYRVARPPVMTDRLG 402  
Db 984 CGWSHLPWPGLGYSWDSGATPSRYPRPS-----VDHTVGT 1021  
QY 403 EVFYF-----GRPAMSLEVERKVFIILCSQNPDIADISHSCL-----HSRKG 442  
Db 1022 EAGHPAFFETSVLPGGQAALGSE-----PLPATAVSCLHFWYMGFFAIFYKG 1071  
QY 443 -LRVLL 447  
Db 1072 ELRVLL 1077

## RESULT 10

US-10-425-114-40809  
; Sequence 40809, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingtong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 40809  
; LENGTH: 810  
; TYPE: PRT  
; ORGANISM: Schizochytrium aggregatum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3033-046-E6\_FLI.pep  
US-10-425-114-40809

Query Match 3.6%; Score 108; DB 12; Length 810;  
 Best Local Similarity 25.2%; Pred. No. 0.94; Indels 62; Gaps 10;  
 Matches 62; Conservative 25; Mismatches 25; Indels 62; Gaps 10;

QY 179 GAEBEDSGDEGSPTRHSGVQVDDANADSPGSDGEGSTRHSDSQPPADETTVHTDN 238  
 DB 467 GSRGSGGRRRGRATSDGLGHGIDNL-ANSAGAA-BGIDVLGLDVE-----HGLN 515  
 QY 239 VEDDLTLDKESA-CALMYHVQGEMLMRAMCDEDLF-DLIGIPEDVIATSPGGDTDA 296  
 DB 516 LGEDLTLDDGVNALSLEHLVGLDHSVLSVSLGGEDLHDSLSGLEKLVDRDSGGSGRR 575  
 QY 297 SGVVTGSIASAVGAGVEDVYL-----AGALEAQNAG-----330  
 DB 576 SGSRSSSGRRRGGGAASDLHLGIDDLTNSASAESVDVLGLDVEHGLNLGENTLTD 635  
 QY 331 -----EYLETISDEV-DGAG-----LPPASRRPVVGEFWDGQPR 368  
 DB 636 GVNALGLEHLVHVLNHSVLSVSLGGEDLHDSLSGLEKLVDRDSGGSGRR-TWGGRR 694  
 QY 369 HERPTT 374  
 DB 695 RGRATS 700

## RESULT 11

US-10-282-122A-64713  
 ; Sequence 64713, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITPA 034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 64713  
 ; LENGTH: 953  
 ; TYPE: PRT

ORGANISM: Mycobacterium tuberculosis  
 US-10-282-122A-64713

Query Match 3.6%; Score 108; DB 12; Length 953;  
 Best Local Similarity 22.0%; Pred. No. 1.2; Indels 160; Gaps 24;  
 Matches 106; Conservative 50; Mismatches 166; Indels 160; Gaps 24;

QY 83 TTSRR-----LRLPQ-----EYEVVQGINCRWRLLCAEVEKWCVHARTLHSGSSLWE 135  
 DB 34 TTSRRVLDALTALDGRVSAHSTVDRDVAVRDLA-----THLETAGVL-- 79  
 QY 136 ILYQHSVLEKRRRPRPPFVGENSDSEED-----HPAF-CDVPVTTQAESEDSDE 188  
 DB 80 ---AASVHAPEASEEPESRLMLETQETRNADVPHYMPLEFVAPQPIPEPLADDEDV-DD 135  
 QY 189 GPSTRHSASGVQVDDANADSPGSDGEGPSTRH-----SDSQP-----226  
 DB 136 GPD-----YVADDSADDEGQLDRPARRRRRRRRRRRRRRRRRRRRKSSAS- 187  
 QY 227 PPA-----DETTVHTDNNVEDDLITLD-----KESACALMYH 257  
 DB 188 PRAQOFTSADAAETDDGDRDSEDTAGDNGEDENGSLAENRRRRRRRRRRRRKSSAS- 242  
 QY 258 VQGEMLMRAMCDEDLFDLLGIPEDVIATSPGCGDTSAGVVTGSLAASAVGAGVEDV 317  
 DB 243 -GDDNDAALGSLPDD-----PPNTVHERVFRAGDKAGNSQDGG-----281  
 QY 318 YLGALEAQNAGEXYVLEISDEVDGAGLPPASRRRPV---GEFLWDDGFRH--BRP 372  
 DB 282 --SGSTEIKGIDGSTLEAKQRDRGR---DAGRRRPPVLSEAEFL---ARREAVRV 332  
 QY 373 TTRIRHKLRSAYYVARPV-----WITDLGLVEVYFGRPAMSLVERKVFIL 423  
 DB 333 MVDRVR-----TEPLPGTRYTQIAVLD--GIVVEHFTVSAASLVGNIVILG 381  
 QY 424 CSQNPFLADISHSLSRKGLRVLLPKDDNNTGPDVNNLLAAVLRSFASGLVIVSLRSGI 483  
 DB 382 IVQNVLPSEAAAFVDIGRNGVL-----YAGEVNWDAAGLGG-ADRKTQALKPGD 432  
 QY 484 YV 485  
 DB 433 YV 434

## RESULT 12

US-09-759-130B-342  
 ; Sequence 342, Application US/09759130B  
 ; Publication No. US2003002279A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Millennium Pharmaceuticals, Inc.  
 ; APPLICANT: McCarthy, Sean A  
 ; APPLICANT: Fraser, Christopher C  
 ; APPLICANT: Sharp, John D  
 ; APPLICANT: Barnes, Thomas S  
 ; APPLICANT: Kirst, Susan J  
 ; APPLICANT: Mackay, Charles R  
 ; APPLICANT: Myers, Paul S  
 ; APPLICANT: Leiby, Kevin R  
 ; APPLICANT: Wrighton, Nicolas  
 ; APPLICANT: Goodearl, Andrew  
 ; APPLICANT: Holtzman, Douglas A  
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING  
 ; PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER  
 ; TITLE OF INVENTION: USES  
 ; FILE REFERENCE: MPI00-5350NMIM  
 ; CURRENT APPLICATION NUMBER: US/09/759,130B  
 ; CURRENT FILING DATE: 2002-09-16  
 ; PRIOR APPLICATION NUMBER: US 09/479,249  
 ; PRIOR FILING DATE: 2000-01-07  
 ; PRIOR APPLICATION NUMBER: US 09/559,497  
 ; PRIOR FILING DATE: 2000-04-27  
 ; PRIOR APPLICATION NUMBER: US 09/578,063  
 ; PRIOR FILING DATE: 2000-05-24



; PRIOR APPLICATION NUMBER: US 09/333,159  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: US 09/596,194  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/342,364  
; PRIOR FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: US 09/508,452  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/393,996  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: US 09/602,871  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: US 09/420,707  
; PRIOR FILING DATE: 1999-10-19  
; NUMBER OF SEQ ID NOS: 460  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 342  
; LENGTH: 883  
; TYPE: PRT  
; ORGANISM: Mus sp.  
US-09-759-130B-342

Query Match 3.6%; Score 107; DB 10; Length 883;  
Best Local Similarity 24.0%; Pred. No. 1.3;  
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;  
QY 151 PRPFFVGEN-----SDSSEEDHPA-----FCDVPTQTGAESDSC-----DE 188  
DB 423 PRTPLESETQSIAPPTESSEEGVALEEEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482  
QY 189 GPSTRHSASGVQVDDANADSPGSGDEGSTRHSQPPPADETTVHTDNVEDDLTLDDK 248  
DB 483 GSETEHSLSQVSPPAQAVQLDASPSGPPR-----FRGPPAE-----TLLP- 524  
QY 249 ESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASGVVTEGSIAS 308  
DB 525 -----PREMSATSTPGG-----AR 538  
QY 309 AVGAGVEDVYLAGALEAQNVAAGEYVLEISDEEVDGAGLPPASRRRPVVGFLWDD-GPR 367  
DB 539 EVGGETGSPELSGVPRESEAGSSLE-----DGPSLLPAT-----WAPVGR 581  
QY 368 RHERPTTRIRHKLRSAYRYVARPPVMTDRL-----GVEVYFGRPAMSLVERKVFILC 424  
DB 582 ELETSPSEKSGRTVL--AGTSVQAQVLPDTSASHGVAV-----APSSG-----DC 626  
QY 425 SONPLADISHSLHRSKGLRVL-LP 448  
DB 627 IPSPCHN-GGTCLKEEKGFRCLCLP 650

RESULT 13  
US-10-189-123-72  
; Sequence 72, Application US/10189123  
; Publication No. US20030082586A1  
; GENERAL INFORMATION:  
; APPLICANT: KIRST, Susan J.  
; APPLICANT: HOLTZMAN, Douglas A.  
; APPLICANT: FRASER, Christopher C.  
; APPLICANT: SHARP, John D.  
; APPLICANT: BARNES, Thomas S.  
; TITLE OF INVENTION: ANTIBODIES HAVING DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER  
; FILE REFERENCE: 10147-11U3  
; CURRENT APPLICATION NUMBER: US/10/189,123  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: US 09/596,194  
; PRIOR FILING DATE: 2000-06-16  
; PRIOR APPLICATION NUMBER: US 09/342,364  
; PRIOR FILING DATE: 1999-06-29  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 72  
; LENGTH: 883

; TYPE: PRT  
; ORGANISM: Mus sp.  
US-10-189-123-72  
Query Match 3.6%; Score 107; DB 14; Length 883;  
Best Local Similarity 24.0%; Pred. No. 1.3;  
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;  
QY 151 PRPFFVGEN-----SDSSEEDHPA-----FCDVPTQTGAESDSC-----DE 188  
DB 423 PRTPLESETQSIAPPTESSEEGVALEEEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482  
QY 189 GPSTRHSASGVQVDDANADSPGSGDEGSTRHSQPPPADETTVHTDNVEDDLTLDDK 248  
DB 483 GSETEHSLSQVSPPAQAVQLDASPSGPPR-----FRGPPAE-----TLLP- 524  
QY 249 ESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASGVVTEGSIAS 308  
DB 525 -----PREMSATSTPGG-----AR 538  
QY 309 AVGAGVEDVYLAGALEAQNVAAGEYVLEISDEEVDGAGLPPASRRRPVVGFLWDD-GPR 367  
DB 539 EVGGETGSPELSGVPRESEAGSSLE-----DGPSLLPAT-----WAPVGR 581  
QY 368 RHERPTTRIRHKLRSAYRYVARPPVMTDRL-----GVEVYFGRPAMSLVERKVFILC 424  
DB 582 ELETSPSEKSGRTVL--AGTSVQAQVLPDTSASHGVAV-----APSSG-----DC 626  
QY 425 SONPLADISHSLHRSKGLRVL-LP 448  
DB 627 IPSPCHN-GGTCLKEEKGFRCLCLP 650

RESULT 14  
US-10-188-495-72  
; Sequence 72, Application US/10188495  
; Publication No. US2003017573A1  
; GENERAL INFORMATION:  
; APPLICANT: KIRST, Susan J.  
; APPLICANT: HOLTZMAN, Douglas A.  
; APPLICANT: FRASER, Christopher C.  
; APPLICANT: SHARP, John D.  
; APPLICANT: BARNES, Thomas S.  
; TITLE OF INVENTION: POLYPEPTIDES HAVING DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER  
; FILE REFERENCE: 10147-11U2  
; CURRENT APPLICATION NUMBER: US/10/188,495  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: US 09/596,194  
; PRIOR FILING DATE: 2000-06-16  
; PRIOR APPLICATION NUMBER: US 09/342,364  
; PRIOR FILING DATE: 1999-06-29  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 72  
; LENGTH: 883  
; TYPE: PRT  
; ORGANISM: Mus sp.  
US-10-188-495-72

Query Match 3.6%; Score 107; DB 14; Length 883;  
Best Local Similarity 24.0%; Pred. No. 1.3;  
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;  
QY 151 PRPFFVGEN-----SDSSEEDHPA-----FCDVPTQTGAESDSC-----DE 188  
DB 423 PRTPLESETQSIAPPTESSEEGVALEEEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482  
QY 189 GPSTRHSASGVQVDDANADSPGSGDEGSTRHSQPPPADETTVHTDNVEDDLTLDDK 248  
DB 483 GSETEHSLSQVSPPAQAVQLDASPSGPPR-----FRGPPAE-----TLLP- 524  
QY 249 ESACALMYHVGQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDASGVVTEGSIAS 308

Db 525 -----PRENSATSTPGG-----AR 538  
Qy 309 AVGAGVEDVLAGALEAONVAGEVLEISDEEVDGAGLPPASRRPVPVGEFLWDD-GPR 367  
Db 539 EVGGETSGPGLGVPRESEAGSSLE-----DGPSLLPAT-----WAPVGPR 581  
Qy 368 RHERPTTTRIRHRKLRGAYRVARPPVMTDRL---GVEVFYFGRPAMSLVERKVFILC 424  
Db 582 ELETSPSEKSGRTVL--ACTSVQAQVLPDTSASHGGVAV-----APSSG-----DC 626  
Qy 425 SONPLADISHSLHRKGLRVL-LP 448  
Db 627 IPSCHN-GGTCLBEKEGFRCLCLP 650

RESULT 15  
US-10-425-114-45409  
; Sequence 45409, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhou, Jingdong  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 45409  
; LENGTH: 1134  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700152164\_FLI.pep  
US-10-425-114-45409

Query Match 3.6%; Score 107; DB 12; Length 1134;  
Best Local Similarity 21.3%; Pred. No. 2;  
Matches 90; Conservative 51; Mismatches 150; Indels 132; Gaps 20;  
Qy 121 VHARTHLHSGSSSLWEI-----LYQHSVRLEKRRPRPPFVGE-----NSD 161  
Db 706 VKAKEHMDAYFSMMQIMSTNQKLSRVRFMLRDSIDLRRNKQORRRKVGPKIEVHRD 765  
Qy 162 SSEEDH-----PACDVPVTQCAESESDGEGPSTRHSASGVQ-PVDDANADSPG 211  
Db 766 AAQERHAQSSRLGEGPAVSSVP---RRHPMDYGRGPSASASSSQSGSIRGNPHPSRG 822  
Qy 212 SGDEGPSTRHSDSQPPADETTVHTDNVEDDLTLDDKESACALMYHVQEMDMLRAMCD 271  
Db 823 SQDIRHDERHQ-----FDNRVTLPQRV-----KDEAITLGPQGLARGMSIR---- 865  
Qy 272 EDLPDLGIGEDVIATQPGDGTASGVVTEGSIASAVGAGVEDVYLAGALEAQNVAQE 331  
Db 866 -----GQP-----PVSNTIPEIVDHRIRIVSSNG-----YNSAAD 896  
Qy 332 YVLEISDEE-----VDDGAGLPASRRPVPVGEFLWDDGPRRHERPTTTRIRHR----- 380  
Db 897 WTSSSGREDSNRLPDRTSGRIPASSQSAVTS-----QRPASQEGRSRSKSYSE 945  
Qy 381 -KLRS-----AYY--RVARPPVMTDRLGVEVFY-----FGPAMSLVERK 419  
Db 946 DELREKSVLTIREYISAKDEKVVLCIEELNAPFYPFLVSLWVNDSPFERKOMERELLAK 1005  
Qy 420 VFI--LCSONPLADISHSLHRK---GLRVLLPKPDN-NTGPGDVNLLAAVLRSFASGL 474  
Db 1006 LLYSLCSGR-----HNLSSKQLSDGLSNVLSLEDNLSDAPEYIGRLRLARFVES 1059  
Qy 475 VIV 477

Db 1060 ILL 1062  
Search completed: June 9, 2004, 09:08:40  
Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 09:01:46 ; Search time 20 seconds  
(without alignments)  
1128.157 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRLTWISFIVGALDSD.....QFDMVPLVILKRLSRVTCDD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 267159 seqs, 39793891 residues

Total number of hits satisfying chosen parameters: 267159

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New.\*

- 1: /cgn2\_6/prodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/prodata/1/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/prodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/prodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/prodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/prodata/1/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/prodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	4.4	364	US-10-779-597-113	Sequence 113, App
2	124	4.1	361	US-10-779-597-105	Sequence 105, App
3	121.5	4.1	2517	US-10-021-698A-122	Sequence 122, App
4	117	3.9	1261	US-60-550-051-213	Sequence 213, App
5	113	3.8	1007	US-10-886-977-24	Sequence 24, Appl
6	112	3.7	327	US-60-556-841-7793	Sequence 7793, App
7	106.5	3.6	686	PCT-US04-07412-706	Sequence 706, App
8	104.5	3.5	503	US-10-796-307-549	Sequence 549, App
9	104.5	3.5	503	US-10-796-307-556	Sequence 556, App
10	104.5	3.5	503	US-60-563-440-1193	Sequence 1193, App
11	104.5	3.5	516	US-10-796-307-552	Sequence 552, App
12	104.5	3.5	516	US-60-563-440-1190	Sequence 1190, App
13	104.5	3.5	1161	US-10-784-004-389	Sequence 389, App
14	104.5	3.5	1161	US-10-784-004-936	Sequence 936, App
15	104	3.5	213	US-60-565-632-9051	Sequence 9051, App
16	104	3.5	323	PCT-US04-05654-1400	Sequence 1400, App
17	103.5	3.5	322	PCT-US04-05654-1401	Sequence 1401, App
18	103.5	3.5	322	US-60-563-282-269	Sequence 269, App
19	103	3.4	357	US-60-556-841-9873	Sequence 9873, App
20	101.5	3.4	2576	PCT-US04-11912-160	Sequence 160, App
21	101.5	3.4	2576	PCT-US04-12049-160	Sequence 160, App
22	100.5	3.4	1177	US-10-796-280-1077	Sequence 1077, App
23	100.5	3.4	1922	US-10-796-280-1080	Sequence 1080, App
24	100.5	3.4	1924	US-10-796-280-1081	Sequence 1081, App
25	100.5	3.4	1966	US-10-828-868-5	Sequence 5, Appl
26	100.5	3.4	1966	US-10-828-868-6	Sequence 6, Appl

27	100.5	3.4	1966	6	US-10-829-000-5	Sequence 5, Appl
28	100.5	3.4	1966	6	US-10-829-000-6	Sequence 6, Appl
29	100.5	3.4	1977	6	US-10-828-868-7	Sequence 7, Appl
30	100.5	3.4	1977	6	US-10-829-000-7	Sequence 7, Appl
31	100.5	3.4	1979	6	US-10-796-280-1079	Sequence 1079, App
32	100	3.3	414	7	US-60-563-440-1194	Sequence 1194, App
33	99.5	3.3	336	1	PCT-US04-05654-1399	Sequence 1399, App
34	99.5	3.3	336	7	US-60-563-282-267	Sequence 267, App
35	99.5	3.3	426	6	US-10-796-280-769	Sequence 769, App
36	99.5	3.3	426	7	US-60-563-440-1015	Sequence 1015, App
37	99.5	3.3	2677	6	US-10-451-566-22	Sequence 22, Appl
38	98.5	3.3	972	6	US-10-630-629-3	Sequence 3, Appl
39	98.5	3.3	1244	7	US-60-556-841-9112	Sequence 9112, App
40	98	3.3	1135	5	US-09-744-794C-16	Sequence 16, Appl
41	98	3.3	1135	5	US-09-744-794D-16	Sequence 16, Appl
42	98	3.3	2468	6	US-10-489-740-216	Sequence 216, App
43	98	3.3	2468	7	US-60-556-903-230	Sequence 230, App
44	97	3.2	595	6	US-10-100-663-6602	Sequence 6602, App
45	96	3.2	991	1	PCT-US04-12717-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1  
US-10-779-597-113  
; Sequence 113, Application US/10779597  
; GENERAL INFORMATION:  
; APPLICANT: Oregon Health & Science University  
; APPLICANT: Wong, Scott W.  
; APPLICANT: Axelsen, Michael K.  
; APPLICANT: Hansen, Scott G.  
; TITLE OF INVENTION: JAPANESE MACAQUE HERPESVIRUS NUCLEIC ACID SEQUENCES AND THEIR USE  
; FILE REFERENCE: 178-67426  
; CURRENT APPLICATION NUMBER: US/10/779,597  
; CURRENT FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: 10/276,524  
; PRIOR FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: PCT/US01/16274  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/205,652  
; PRIOR FILING DATE: 2000-05-18  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 113  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Japanese Macaque Herpesvirus  
US-10-779-597-113

Query Match	4.4%	Score 133;	DB 6;	Length 364;
Best Local Similarity	23.4%	Pred. No. 0.012;		
Matches	82;	Conservative	48;	Mismatches 135;
				Indels 86;
				Gaps 18;
QY	246	LKESACALMYHVQCEMDMLMRACDEDLFDLLGIPEDVIATSQPGDITDASGVVTEGSI	305	
DB	57	LDRDLQGRQHNVSDCKRLLVLRNAGPE---QDDTRATTNSG-----	99	
QY	306	AASVAGVGVYVYLALEAQNVAQVYVLEISDEVDGGA-----GLPPASRRPVVGE	359	
DB	100	-----GERFFYLKPAVDPLCYA--CILDSSHSETVINYLEAACVHGLEP-----	140	
QY	360	FLWDDGPRRHERPTTRIRHRKLSAYYVARP--PVMITD--ELGVEVYFGRFAMSLE	415	
DB	141	--WTFPLQPAEASGAARSVYAAARLATAAPHPQITFFWRLRQVFFG-----SLV	194	
QY	416	VE-----RKVFILCSQNPLADISHSCHSRKGRVLRLLPKPDDNN--TGPQVNLAAVLR	468	
DB	195	AEHTGVDRRGVRLHKRQDPKA--GHACYG-TAFKMWLPTPHENGPLTPEQRETCVBIIN	251	
QY	469	SFASGLIVLSRSGIYVKNLCKSTVLYHGN---NPKKSGVICGLSSRAVLDFVFN---	521	
DB	252	YCEEGIFLHGNELGYVDNTRHRTLSAGNDAGNHAQRF-----VRSCKAFQIFYVWGL	306	

Qy	522	----	AQRIGQE-HIKKTTVIGGDPTSAQFOWPLVIKRLRSVTCD	567
Db	307	LRLKLSPVGDFPFINAVTLYLGRGSRK-POVPIL-----VICD	350	

## RESULT 2

```

US-10-779-597-105
/ Sequence 105, Application US/10779597
/ GENERAL INFORMATION:
/ APPLICANT: Oregon Health & Science University
/ APPLICANT: Wong, Scott W.
/ APPLICANT: Axthelm, Michael K.
/ APPLICANT: Hansen, Scott G.
/ TITLE OF INVENTION: JAPANESE MACAQUE HERPESVIRUS NUCLEIC ACID SEQUENCES AND THEIR USE
/ FILE REFERENCE: 178-67426
/ CURRENT APPLICATION NUMBER: US/10/779,597
/ CURRENT FILING DATE: 2004-02-12
/ PRIOR APPLICATION NUMBER: 10/276,524
/ PRIOR FILING DATE: 2002-11-13
/ PRIOR APPLICATION NUMBER: PCT/US01/16274
/ PRIOR FILING DATE: 2001-05-17
/ PRIOR APPLICATION NUMBER: 60/205,652
/ PRIOR FILING DATE: 2000-05-18
/ NUMBER OF SEQ ID NOS: 172
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 105
/ LENGTH: 361
/ TYPE: PRT
/ ORGANISM: Japanese Macaque Herpesvirus
/ US-10-779-597-105

```

Query Match		4.1%;	Score 124; DB 6; Length 361;
Best Local Similarity		22.5%; Pred. No.	0.06;
Matches	47; Conservative	31; Mismatches	91; Indels 40; Gaps 8;
QY	383 RSAYRYRVAR-----PPVMTD--RLGVSVFVFGRPAMLSVERKVFILCSQNPLADIS	433	:::::
Dd	155 RSYARAHRLGVPPEAPLPHQLIVPWRLRIQUFFVGVLALDHTSQDRRGVRLEPRFPDPGP	214	:::::
QY	434 HSLCHSRKLGLRVLLPKPDNDNNTGPSVDNNLLAAVLRSFASGLVIVSRLSGIYYKNLCSTV	493	: :::
Dd	215 HLCPFG-TGFVTWPFSPDGKLTPOQTQINTMLVTYNEGIIYVHGNETGYVDNRRETL	273	: :::
QY	494 LYHGNPNPKPGVICG---LSSRAVLDFVNVAQYRIOG-----HEHIKKITVF	538	:::
Dd	274 YAGAND-----CNGDIQREWFMFLSKOOIFYFNMGFKRLARSVPESHACPNCATLY	325	:::
QY	539 IGGDPTSAEQDMVPLVKLRLSRTCCDD	567	:
Dd	326 LSQLQP-GAQESFPQPISV- ----VVQCDD	347	:  :::

### RESULT 3

US-10-021-698A-122  
 ; Sequence 122, Application US/10021698A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KEITH, TIM  
 ; APPLICANT: LITTLE, RANDALL  
 ; APPLICANT: VAN EERDEWEGH, PAUL  
 ; APPLICANT: DUPUIS, JOSSE  
 ; APPLICANT: DEL MASTRO, RICHARD  
 ; APPLICANT: SIMON, JASON  
 ; APPLICANT: ALLEN, KRISTINA  
 ; APPLICANT: PANDIT, SUNIL  
 ; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO  
 ; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY  
 ; FILE REFERENCE: 2976-404AUS1  
 ; CURRENT APPLICATION NUMBER: US/10/021, 698A  
 ; CURRENT FILING DATE: 2001-10-22  
 ; PRIOR APPLICATION NUMBER: 60/211, 749  
 ; PRIOR FILING DATE: 2000-06-14  
 ; NUMBER OF SEQ ID NOS: 6160

```

; SOFTWARE: Patentin 2.1
; SEQ ID NO 122
; LENGTH: 2517
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-698A-122

```

Query Match	4.1%;	Score 121.5;	DB 6;	Length 2517;
Best Local Similarity	21.3%;	Pred. No. 1.4;		
Matches	81;	Conservative 48;	Mismatches 132;	Indels 119; Gaps 17;
QY	105	RRWELLCABV---KCCWCVHARTHLHSGSLWEILYQHSVRLEKRRPRPRPFVGENSDS	162	;
Db	632	RWSATARWGVSTVQCCKNFYNYKKRQWDLILQHKLMKERNARRKKKAPAPAAAS	691	;
QY	163	SEEDHPAFCDVPVYQ-----TGAASE-----DSGDE-----GPSTRHAS	197	;
Db	692	EZAFAFP-----PVVEDEEMEAASGVSGNEEMVEEAALHASGNEVPRGCGSPATVNSS	746	;
QY	198	GVQ-----PVDDANADSPGSDGEGPSTRHSDSQP-----PADETT---	233	;
Db	747	DTESIFSPTEAAKDTGQNGPKPATLGADGPPGPTPPRRTSRAPIEPTTPASEATGAP	806	;
QY	234	-----YHTDNVEDDLTLLOKESACALMYHVQEWMDMLMAMCDEDLFDLL	278	;
Db	807	TPPPAPPSAPPVPVPKEEK-----ETAAAPPVEEGE-----QKPPAAEELAVDT	855	;
QY	279	GIPEDVIAT-----QPGGDTAGVVVTGSGTAASAVGAGVEDVYLAGALEAONVA	329	;
Db	856	GKAAEFPVKSECTEEAEEGPAKGRDAEAAETAGALKAEKKEGG-----SGRATTAKSS	909	;
QY	330	G-----EYVLETSDEBVDGAG-----LPPASRRRPVVGVEFLWDGPPRRHERP-TTTRI	377	;
Db	910	GAPQDSSSATCSADENVDAEGDKKRLSLPRPSLLTPTGD-----PRNASQKPLDL	963	;
QY	378	RHRKLSAYYRVARPVPMIT	397	;
Db	964	KOLKORAA-----AIPPIQVT	979	;

## RESULT 4

```

US-60-550-051-213
; Sequence 213, Application US/60550051
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001508
; CURRENT APPLICATION NUMBER: US/60/550,051
; CURRENT FILING DATE: 2004-03-05
; NUMBER OF SEQ ID NOS: 23014
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-550-051-213

```

	Query Match	3.9%; Score 117; DB 7; Length 1261;
	Best Local Similarity	19.7%; Pred. No. 1.2;
	Matches 122; Conservative	55; Mismatches 187; Indels 256; Gaps 29;
QY	24 LVKKLDRSTGTFPLAARNDVPLDLSLOFFIDFK-----RECL----SKGLHPRDLIG 72	: : : : : :
Db	221 LYAW-DNRRTFRKSLGNDETDEKKKFLGFFKNVKSNKGCLLTTPNSPMHSRLTL 279	: : : : : :
QY	73 SPIAFAGKICTTS-----RRLRLPGEYEYWGQINGCRRWRLLCAEVEKCWCVCUART 125	: : : : : :
Db	280 GPSILSLSGISGVSVKSEMKRRAPPFGSGPPVQD-----KASE 318	: : : : : :
QY	126 HLHGSSLWEILYCHSVRLKEKHRR-----PREPFVGENSEDSSEEDHPA-FCDDVP 174	: : : : : :
Db	319 KVSISGS-----QIDLOKKRRAPADPPPPPSPFPINPTDEKNEKRKSMTWSLP 369	: : : : : :

QY 175 VTOTGAESDSESGDEPSTRHSASGVQVDDANADSPGSDGEGSTRHSQPPADETTV 234  
Db 370 L---GSGSHCSPOGAPQVLSAEETVSGSCFASDTEGWSFSD-----IV 417  
QY 235 HTDVEDDLTLDK-----ESACALMYHG-----259  
Db 418 SLDSQDSQSMYKDKWATDQEDCSQDLAGTDLGPKSPLWKNKNGSENSHLRTEKAVTAS 477  
QY 260 -QEMDMLMR-----AMCDEDLPLIGIPEDVIATSQ-----PG 291  
Db 478 NDEEDLIIAGEPRKTLAEDELEEM-----EDSVETDTSLTSIHGASHNCFQDAMIHP 533  
QY 292 GDTDASGVWTEGSAASAVAGVEDVYLGALEAQNVAAGEVYLEISDEEYDDGAGLPPAS 351  
Db 534 GDTDAIPVTFIGVSDPVDPSGLFS-----NRNNNAGSF-----DSEGV--AS 574  
QY 352 RR---RPVVGFEFLWDDGPRHERPTTRIRHKLRSAYYVARPPVMTDRLGVVEFYFG 408  
Db 575 RRASLAPLOAE-----HSQP-----HEKARE-----EV-----597  
QY 409 RPAMSLVERKVFILCSQNLADISHSLHS-RKGLRVLLPKPDNDNTGPDVNNLLAAVL 467  
Db 598 -PAL-----HPASHDIGIRVAL-----SNISKDGNLMETAPRV 631  
QY 468 RSFASGL-----VIVLSRSGY-----VNNCKSTVLYH--496  
Db 632 TSFASNLHTDNLNAKVKDKVYGCADGERTQATERVNSQPVNEKDSNDKNAALAPTSHQOR 691  
QY 497 GNPPEKFGVICGLSSRAVL 516  
Db 692 GQNFKSYRLKHGLTYYKII 711

## RESULT 5

US-10-486-977-24

Sequence 24, Application US/10486977

## GENERAL INFORMATION:

APPLICANT: INCYTE CORPORATION; BAROSSO, Ines;  
APPLICANT: BAUGHN, Mariah R.; BECHA, Shanya D.;  
APPLICANT: BLAKE, Julie J.; BOROWSKY, Mark L.;  
APPLICANT: BURFORD, Neil; DUGGAN, Brendan M.;  
APPLICANT: ELLIOTT, Vicki S.; EMERLING, Brooke M.;  
APPLICANT: FORSYTHE, Ian J.; GIETZEN, Kimberly J.;  
APPLICANT: GORVAD, Ann E.; GRIFFIN, Jennifer A.;  
APPLICANT: HAPALIA, April J.A.; HONCHELL, Cynthia D.;  
APPLICANT: ISON, Craig H.; KHAN, Farrah A.;  
APPLICANT: LAL, Preeti G.; LEE, Ernestine A.;  
APPLICANT: LEE, Sally; LEE, Soo Yeun;  
APPLICANT: LI, Joana X.; LU, Dying Aina M.;  
APPLICANT: LU, Yan; LEHR-MASON, Patricia M.;  
APPLICANT: NGUYEN, Daniel B.; RAMKUMAR, Jayalaxmi;  
APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;  
APPLICANT: THANGAVELU, Kavitha; THORNTON, Michael B.;  
APPLICANT: TRAN, Uyen K.; CHAWLA, Nazinder K.;  
APPLICANT: WARREN, Bridget A.; XU, Yuming;  
APPLICANT: YAO, Monique G.; YUE, Henry;  
APPLICANT: YUE, Huibin; ZEBARJADIAN, Yeganeh

TITLE OF INVENTION: NUCLEIC-ACID ASSOCIATED PROTEINS

FILE REFERENCE: PF-1146 USN

CURRENT APPLICATION NUMBER: US/10/486,977

PRIOR FILING DATE: 2004-02-17

CURRENT APPLICATION NUMBER: PCT/US02/25829

PRIOR FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: US 60/313,111

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: US 60/314,682

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/314,756

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/315,105

PRIOR FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: US 60/316,751

## RESULT 6

US-60-556-841-7793

Sequence 7793, Application US/60556841

## GENERAL INFORMATION:

APPLICANT: Abad, Mark S.  
TITLE OF INVENTION: Genes and Uses for Plant Improvement  
FILE REFERENCE: 38-21(53450)  
CURRENT APPLICATION NUMBER: US/60/556,841  
CURRENT FILING DATE: 2004-03-25  
NUMBER OF SEQ ID NOS: 12463  
SEQ ID NO 7793  
LENGTH: 327

PRIOR FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: US 60/316,856  
PRIOR FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: US 60/328,185  
PRIOR FILING DATE: 2001-10-05  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PERL Program  
SEQ ID NO 24  
LENGTH: 1007  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No: 3773014CD1  
US-10-486-977-24

Query Match 3.8%; Score 113; DB 6; Length 1007;

Best Local Similarity 19.4%; Fred. No. 1.8; Indels 228; Gaps 25;

Matches 111; Conservative 61; Mismatches 171;

QY 146 KHRPRRR-----PFVGENSDSEEDHPAFCDVPTVQ-----TGAESD-----184  
Db 4 RKQRKQQLISDCGSPASENGDASEEDHPQVCAKCAQFTDPTFLAHQACSTDPVPM 63  
QY 185 ---SGDEPSTRHSASGVQP-----VDANADSPGSDG-----GPSTRHSDSQ-225  
Db 64 VIIGQENPNNSASSEPRPEGHNNPQVMDTEHSPDSSGVPTDPTWGPERRGESG 123  
QY 226 ---PPADETTVHTDNVE-----240  
Db 124 HELVAATGTAAGGGGLILASPKLGATLPPESTAPPAPPPPPPPGVSGLNIPIL 183  
QY 241 DDLTLDK-----ESAC---ALMYHVGQ-----EMDMLRAMCDEDLFDLLG 279  
Db 184 BELRVLQQRQHQMOWTEBQICRQVLLGSLGQTVGAPASPSLPGTGTASSTKPLPLFS 243  
QY 280 IPEDV-----IATSPGGDT-----DASGV-----299  
Db 244 PIKPVQTSKTLASSSSSSSSSGAETPKQAFPHLYHPLGSHQHPFSAGGVGRSHKPTAPS 303  
QY 300 ---VTEGSIASAASVGA---GVEDVYLGALEAQNVAAGEVYLEISD 338  
Db 304 PALPGSTQLIASPHLAPFSTIGLLAAQCLGAARLEATASGLLKPKNGSG---ELSY 359  
QY 339 EVDDGAGLPPASRRRPVVGFLWDDGP-----RRH---ERP-----TTR---RIR 378  
Db 360 GEVMGLEKPGGRHKRCFCAKVFGSDSALQIHLRSHTGERTGKVCNCGNRFTTRGNLVH 419  
QY 379 HRKLSAYYVARPPVMTDRLGVEVFFVGRP-AMSL------VERKVFIL 423  
Db 420 FHRREKYPHYQMNPHVPEHLDVITSSGLPYGMSVPEKABEAAATPGGVERKPLVA 479  
QY 424 CSQNPLADISHSLHSRKG-----LRVLLPK-PDDNNTGPG-DVNLLAAV 466  
Db 480 STTALSATESLTLLSTASGTATAPGLPAFNKFLVMKAVEPKNKADETPPGSEGAISGV 539  
QY 467 LESFAS-----GLVIVLSRSGIYVKNLCKST 492  
Db 540 AESSTATRQVLSKLVTSPLSWALLTNHFKST 570

## RESULT 6

US-60-556-841-7793

Sequence 7793, Application US/60556841

## GENERAL INFORMATION:

APPLICANT: Abad, Mark S.  
TITLE OF INVENTION: Genes and Uses for Plant Improvement  
FILE REFERENCE: 38-21(53450)  
CURRENT APPLICATION NUMBER: US/60/556,841  
CURRENT FILING DATE: 2004-03-25  
NUMBER OF SEQ ID NOS: 12463  
SEQ ID NO 7793  
LENGTH: 327

```

; TYPE: PRT
; ORGANISM: Oryza sativa
US-60-556-841-7793

Query Match
Best Local Similarity 3.7%; Score 112; DB 7; Length 327;
Matches 75; Conservative 42; Mismatches 132; Indels 82; Gaps 13;

QY 253 ALMYHVGQMDMLRAMCDEDLFDLLGIPEDVIATSPQGGDTDAGVVTGSGIAASAVGA 312
DB 53 ALAFHAGQ-----AVAIPIHVDIAPRKEGPI-AVGAQILAEAGAGLFS 95
QY 313 GVEDVILAGALBAQNVAEYVLEISDEVDGAGLPPASRRPV-----VGEFLMD--- 363
DB 96 GVSATMLRQTLSTYTRMGLYDI-LKKWTOENGGLVPLHRKIAAGLIAGGVGAAGVNPAD 154
QY 364 -----DGRPRHERPTTRIRHKLRSAYYRVARPFVMTDRLGVEVEFGRRPMSLE 415
DB 155 VAMVRQADG-----RLPLAERNYRVSVDALGRMARDE-----GVRLWRG---SSLT 200
QY 416 VERKVFILCSQNPDLADISHCSLHRSKGLRLVLLPKPDDNNTGPDVNLAAVLRFSAGLV 475
DB 201 VNRAMIVTASQATVDQAKEALARR-----GFGADGLGTHVVVAFAGLV 246
QY 476 IVSLRSGI-YVKNLCKSTVLYHGNPKKFGVIGLSRAVLVDVFNVAQYRIQGEHIKK 534
DB 247 AAAANPVDVVKTRVNNMKVAPGAPPYSGAIDCAKLT-----VRSEG-----V 290
QY 535 TTVFTIGDPTSABQ--FDWPLVIXLRLSV 563
DB 291 MALYKGFPTVSRQGFPTWLPVTLQVRKV 321
```

```

RESULT 7
PCT-US04-07412-706
; Sequence 706, Application PC/TUS0407412
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-rui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Ghosh, Malabika
; APPLICANT: Xue, Aidong J.
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Dunrui
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Gezhi
; APPLICANT: Wang, Zhiwei
; APPLICANT: Boyle, Bryan J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 822CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US04/07412
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: US 10/389,559
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 60/365,264
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/340,187
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 10/296,115
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
```

```

; PRIOR APPLICATION NUMBER: US 10/275,027
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920
; SOFTWARE: PCT-FL_genes Version 6.0
; SEQ ID NO 706
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-07412-706

Query Match
Best Local Similarity 3.6%; Score 106.5; DB 1; Length 686;
Matches 61; Conservative 17; Mismatches 86; Indels 71; Gaps 11;

QY 179 GAESDSQDEG-----PSTRHSASGVQFVDNADSPSGSDEG----- 216
DB 78 GAHGETEAEAGAEVPGGEEETSGAQVEGA---SPGRGAQGEPRGEAQREPEDSAA 134
QY 217 PSTRHSDSQPPADETTVH-----TDNVEDDLTLDDKESACALMYHVQEMD 263
DB 135 PERQEAEORPEVPEGSASGEAGDSVDAEGPLGDNIEAGPGAGDSVEAEG---RVGDSVD 191
QY 264 MLMRA-----MDE-----DLFDLLGIPEDVIATSPQGGDT-DASGVVTEGSI 305
DB 192 AEGPAGDSVDAEGPLGDNIAEGPAGDSVDAEGPAGDSVDAEGPAGDSVDAEG 251
QY 306 AASAVGAGVEDVYLAG-ALEAQNVAEYVLEISDEVDGAGLPPASRRRPVGE 359
DB 252 AGDPAGDGEVAGVPAGDSVDAEGPAGDSM-----DAEG--PAGRRARRVSGE 295
```

```

RESULT 8
US-10-796-307-549
; Sequence 549, Application US/10796307
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001509
; CURRENT APPLICATION NUMBER: US/10/796,307
; CURRENT FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 44201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 549
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-796-307-549

Query Match
Best Local Similarity 3.5%; Score 104.5; DB 6; Length 503;
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 26;

QY 11 EFIVGALDSKYPVKWLDNRSTGTFLAP---AARNVDVPLDSLQFFIDFKRECLSKGLHP 67
DB 16 EMLIGEISSGCGYGLQWLDEATCTPRVPMKHFARKDLSADAR----IPKAWAVAGRWP 71
QY 68 RDLGSPITAFGKICTTSRRRLRPLGEEYEVVQGINCRWRLLCAEVKECWCVHARTHL 127
DB 72 PSSRGGG-----PPPEAETA-----RAGKNTNFRCAL 99
QY 128 HGSLSLWEILYQHSVRLEKHRRRPRPF-----VGENSDSSEEDHPACDVPV 175
DB 100 RS-----TRRFVMLRDNSGDPADPHKVVYALSRELWCWREGPGTQDTEAEAPAAVPPP- 150
QY 176 TGTGAESDSGDEGSTRHSASGVQFVDNADSPSGSDEGSPSTRHSQPPADETTVH 235
DB 151 -----QGGPFGFLATHAGLQ-----AFG-----PLPAP----- 175
```

```

236 TDNVEDDLTLDKESACALYHVGOEMDMLRACMEDDLFOLLGIPEDVIATSPGGDTD 295
176 -----ADKGDLTLQAVQCSCLADHL-----LTSAGWD-- 204
296 ASGVVTEGSIATAASAVGAGVEDVYLAGALEAQNAGEVYVLEISDEVEDDGAGLP----- 348
205 -----PVPTKAPGGQEGCLPLTGA-----CAG-----GPGLPAGELYGW 238
349 -----PASRRRP-----VVGEFLWDDGPRRHE--RPTTTRIRHRKLRSAIYVRVAPPPVMIT 397
239 AVETTPSGQPQPAALTTEGAAPSPHQAEPLYSPSP-----SACTAVQEPSP---P 286
398 DRUGVEVYFCRPAMSLVERK--VFILCSONPLADISHSCLHSRKGRLRVLLPKPDDNNT 455
287 GALTVTIMYKGRVTLOKVVGHPSCTFLYGPDP-----AVRATDPOQVAFPSPAEL-- 337
456 GPGDVNL--LAAVURSFPASGLIVSVLSRGIYVKNLCKSTVLYHGNNPPKFG--VICGL 510
338 -POKQLRYTEELLRHVAPGLHLRGLPQLWARDMGCKVYWEVGGPPGSPASPTPACLL 396
511 SSRAVLDVEN-----VAQYRIQGHEHIKXTTVFIF-GDPTSA 546
397 PRNCDTPIFPRVFQELVEFRQRGRSPRYTIVLFGQDLSA 440

RESULT 9
US-10-796-307-556
; Sequence 556, Application US/10796307
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01509
; CURRENT APPLICATION NUMBER: US/10/796,307
; CURRENT FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 44201
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 556
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-796-307-556

```

Query Match	3.5%;	Score 104.5;	DB 6;	Length 503;
Best Local Similarity	19.0%;	Pred. No. 3.2;		
Matches 111;	Conservative 61;	Mismatches 205;	Indels 207;	Gaps 26;
QY	11	EFIVGALDSKYPLVKKWLDSTGFLAP--AARDNVIPLDSLQFFIDFKECUSKGLHP	67	
Db	16	EWLGEISSGCGYEGQLWDEARTCFRVWKHFARKDSEADAR---IFKAWAVARGWP	71	
QY	68	RDLLGSPITAFGKICTTSSRLRLRFLGGEYEVVQGINCRMRLLCAEVKECWCVHARTHL	127	
Db	72	PSRRGGG-----PPPEAETAE-----RAGWKTNFCAL	99	
QY	128	HSQSSLWEIYYQHSVRLEKHHRRRRPRP-----VGNSDSSSEEDHPAPCDVPV	175	
Db	100	RS-----TRREVMLRDNSGDPADPHKVYALSRCLCWREGPGCTDQTEAEAPAAVPPP-	150	
QY	176	TQTGAESEDSDGEGPSTRHSASGQVPVDDANADSPGSGDEGPFSTRHSDSQPPADETTVH	235	
Db	151	-----QGGPPGPFIAHTHAGLQ-----APG-----PLPAP-----	175	
QY	236	TDNVEDDLTLDKESACALMYHVQCEWMDMLRAMCDEDLFDLLGIPEDVIATSPGGTDT	295	
Db	176	-----AGDKGDLQLQAQQSCSLADHL-----LTASWGAD--	204	
QY	296	ASGVVVTGSIATAASAVGAGVEDVYLAGALEAQNVAGEYVLEISDEVDGAGLP-----	349	
Db	205	-----PVPTKAPGSGEGGLPLTGA-----CAG-----GPGLPAGELYGW	238	
QY	349	-----PASRRRP-----VVGEFWLDDGPRRHE---RPTTTRIRHRKLRSAYYVAVRPPVMT	397	

```

Db      239 AVETTPSPGPPAALTGEAAAPESPHQAEFYLSPPP-----SACTAVOEPS---P 288
Qy      398 DRLGEVEYFGRPAMSLVERK--VFTLCSONPLADISHSCLHSRKGRLVLFLPKDDNNT 455
Db      287 GALDVTIMYKGRTVLQKVGHVHSPCTFLGGPDP-----AVRATDPQQVAFPSPAEL-- 337
Qy      456 GFGDVNL--LAAVLRSPASGLIVSVLSRSYGIVYKNLCKSTVLYHGNNPPKKPG---VICGL 510
Db      338 -PDQQLRYTEELLRHVAPGLHLERGPQLWARRMGCKKYMEVGPPGSGASPSPACLL 396
Qy      511 SSRADVDFN-----VAQYRIQGHEHIKTTVFVG-GDPTSA 546
Db      397 PRNCDTPIFDPRVFQBLVEFRARQRGSPRYTIYLGFQGDUSA 440

RESULT 10
US-60-563-440-1193
; Sequence 1193, Application US/60563440
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: CHANG, Sheng-Yung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/563,440
; CURRENT FILING DATE: 2004-04-20
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 1193
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-563-440-1193

Query Match          3.5%; Score 104.5; DB 7; Length 503;
Best Local Similarity 19.0%; Pred. No. 3.2;
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 266
```

[illegible]



QY 456 GPGDVL--LAALRSFASGLVIVSLRSGIYVKNLCKSTVLYHGNPPKFG--VICGL 510  
Db 338 -PDQQLAYTEELLRHVAPGLHLELRGPQLWARRMGCKVYVEVGGPGSASPACPCLL 396  
QY 511 SSRVLDFN-----VAQRIQGEHIKKTTFVIG-GDPTSA 546  
Db 397 PRNCPTPIDFRVFFRELVEFRARQRSGSPRYTIIYLGFGQDLA 440

## RESULT 11

US-10-796-307-552  
; Sequence 552, Application US/10796307  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001509  
; CURRENT APPLICATION NUMBER: US/10/796,307  
; CURRENT FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 44201  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 552  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-796-307-552

Query Match 3.5%; Score 104.5; DB 6; Length 516;  
Best Local Similarity 19.0%; Pred. No. 3.3;  
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 26;

QY 11 EFTVGLDSDKYPLVKWLDSTGTFLAP---AARNVPLDLSQFFIDFKRECLSKGLHP 67  
Db 29 EMLLGEISSGCVGEGQLWLDDEARTCFRVWKHFARKDLEADAR----IFKAWAVARGRP 84  
QY 68 RDLGSPITAFGKICTTSRRLRFLPGEYEVVQGINCRRLCAEVKECWCVCVHARTHL 127  
Db 85 PSSRGGG-----PPPEAETAE-----RAGWKTNFRCAL 112  
QY 128 HGSLSLWEILYQHSVLEKHEHRRPRPF-----VGENSDSSEEDHRAFCVVPV 175  
Db 113 RS-----TTRFVNLRDNSGDPADPHKVYALSRELWCWREGPGTDOTEAEAPAVPPP- 163  
QY 176 TOTGAESDSGDEGSTHRSASGVQVDDANADSPGSDGEGPSTRHSOSPAPADETTVH 235  
Db 164 -----QGGPPGPFLAHTAGLQ-----APG-----PLPAP----- 188  
QY 236 TDNVEDDLTLDKESACALMTHVQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGTD 295  
Db 189 -----AGDKGDLQLQAVQQSCLDHL-----LTASWGAD-- 217  
QY 296 ASGVVTEGSIASAAGVAGVEDVYLAGALEAQNVAGEYVLEISDEEVDGAGLP----- 348  
Db 218 -----PVPTKAPGEQEGPLTGA-----CAG-----GPGLPAGELYGW 251  
QY 349 -----PASRRRP---VVGFEFLWDDGPRRHE---RPTTRIRHRKLRGAYRVARPPVMT 397  
Db 252 AVETTPSPGPQPAALTGTGEAAAPESPQAEFYLSPSP-----SACTAVQEPS---P 299  
QY 398 DELGVEVEYFGRPAMSLVERK--VFILCSQNLADISHSCLHSRKGRLVLLPKDDNNT 455  
Db 300 GALDVTIMYKGRVTLQKVGHPSCTFLYGPDP-----AVRATDPOQVAFPSPAEL-- 350  
QY 456 GPGDVL--LAALRSFASGLVIVSLRSGIYVKNLCKSTVLYHGNPPKFG--VICGL 510  
Db 351 -PDQQLAYTEELLRHVAPGLHLELRGPQLWARRMGCKVYVEVGGPGSASPACPCLL 409  
QY 511 SSRVLDFN-----VAQRIQGEHIKKTTFVIG-GDPTSA 546  
Db 410 PRNCPTPIDFRVFFRELVEFRARQRSGSPRYTIIYLGFGQDLA 453

## RESULT 12

US-10-784-004-389  
; Sequence 389, Application US/10784004  
; GENERAL INFORMATION:  
; APPLICANT: Biogen Idec  
; TITLE OF INVENTION: Surrogate Markers of Pain  
; FILE REFERENCE: 08201.6029-00000  
; CURRENT APPLICATION NUMBER: US/10/784,004

US-60-563-440-1190  
; Sequence 1190, Application US/60563440  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: CHANG, Sheng-tung  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001470  
; CURRENT APPLICATION NUMBER: US/60/563,440  
; CURRENT FILING DATE: 2004-04-20  
; NUMBER OF SEQ ID NOS: 47859  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1190  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-563-440-1190

Query Match 3.5%; Score 104.5; DB 7; Length 516;  
Best Local Similarity 19.0%; Pred. No. 3.3;  
Matches 111; Conservative 61; Mismatches 205; Indels 207; Gaps 26;

QY 11 EFTVGLDSDKYPLVKWLDSTGTFLAP---AARNVPLDLSQFFIDFKRECLSKGLHP 67  
Db 29 EMLLGEISSGCVGEGQLWLDDEARTCFRVWKHFARKDLEADAR----IFKAWAVARGRP 84  
QY 68 RDLGSPITAFGKICTTSRRLRFLPGEYEVVQGINCRRLCAEVKECWCVCVHARTHL 127  
Db 85 PSSRGGG-----PPPEAETAE-----RAGWKTNFRCAL 112  
QY 128 HGSLSLWEILYQHSVLEKHEHRRPRPF-----VGENSDSSEEDHRAFCVVPV 175  
Db 113 RS-----TTRFVNLRDNSGDPADPHKVYALSRELWCWREGPGTDOTEAEAPAVPPP- 163  
QY 176 TOTGAESDSGDEGSTHRSASGVQVDDANADSPGSDGEGPSTRHSOSPAPADETTVH 235  
Db 164 -----QGGPPGPFLAHTAGLQ-----APG-----PLPAP----- 188  
QY 236 TDNVEDDLTLDKESACALMTHVQEMDMLRAMCDEDLFDLLGIPEDVIATSPQGGTD 295  
Db 189 -----AGDKGDLQLQAVQQSCLDHL-----LTASWGAD-- 217  
QY 296 ASGVVTEGSIASAAGVAGVEDVYLAGALEAQNVAGEYVLEISDEEVDGAGLP----- 348  
Db 218 -----PVPTKAPGEQEGPLTGA-----CAG-----GPGLPAGELYGW 251  
QY 349 -----PASRRRP---VVGFEFLWDDGPRRHE---RPTTRIRHRKLRGAYRVARPPVMT 397  
Db 252 AVETTPSPGPQPAALTGTGEAAAPESPQAEFYLSPSP-----SACTAVQEPS---P 299  
QY 398 DELGVEVEYFGRPAMSLVERK--VFILCSQNLADISHSCLHSRKGRLVLLPKDDNNT 455  
Db 300 GALDVTIMYKGRVTLQKVGHPSCTFLYGPDP-----AVRATDPOQVAFPSPAEL-- 350  
QY 456 GPGDVL--LAALRSFASGLVIVSLRSGIYVKNLCKSTVLYHGNPPKFG--VICGL 510  
Db 351 -PDQQLAYTEELLRHVAPGLHLELRGPQLWARRMGCKVYVEVGGPGSASPACPCLL 409  
QY 511 SSRVLDFN-----VAQRIQGEHIKKTTFVIG-GDPTSA 546  
Db 410 PRNCPTPIDFRVFFRELVEFRARQRSGSPRYTIIYLGFGQDLA 453

## RESULT 13

US-10-784-004-389  
; Sequence 389, Application US/10784004  
; GENERAL INFORMATION:  
; APPLICANT: Biogen Idec  
; TITLE OF INVENTION: Surrogate Markers of Pain  
; FILE REFERENCE: 08201.6029-00000  
; CURRENT APPLICATION NUMBER: US/10/784,004



QY 255 MYHVQEMDMMLRAMCDEDLFDLLGIPEDVIATSPGGDTDASG-----VVTEGSIA 306  
Db 136 -----SPTDSPPEPATDSSASPTDSPDSTTASSSGSSTLLSDSTGATPESTKGSAA 190  
QY 307 ASAVG 311  
Db 191 SLFTG 195

Search completed: June 9, 2004, 09:07:37  
Job time : 22 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	121.5	4.1	1166	2	T15628
2	117.5	3.9	1567	2	hypothetical protein antigen containing
3	117	3.9	1316	2	KIA0633 protein - tegument protein 6
4	115.5	3.9	3436	2	S55659
5	113.5	3.8	1005	2	C71513
6	111	3.7	1216	2	hypothetical protein apical endosomal p
7	109	3.6	1664	2	probable retroelasm
8	109	3.6	4717	2	F84485
9	108	3.6	953	2	T41581
10	107	3.6	836	2	B70681
11	107	3.6	883	2	B84417
12	107	3.6	974	2	S57653
13	106.5	3.6	347	2	A40580
14	106.5	3.6	1196	2	S35994
15	105.5	3.5	827	2	T04789
16	105	3.5	813	2	A12651
17	105	3.5	1275	2	H97433
18	105	3.5	1892	2	T18314
19	104.5	3.5	1161	2	I59311
20	104	3.5	852	2	A38459
21	103.5	3.5	273	1	EDBE12
22	103.5	3.5	284	2	T42608
23	103	3.4	357	2	P70898
24	102	3.4	950	2	S27473
25	102.5	3.4	984	2	C84781
26	102.5	3.4	1506	2	A96808
27	102.5	3.4	2124	2	A28452
28	102	3.4	690	2	H84309
29	102	3.4	815	2	H96494
30	102	3.4	815	2	protein p7f22.2 [i



Query Match 3.9%; Score 115.5; DB 2; Length 3436;  
Best Local Similarity 21.4%; Pred. No. 9.8;  
Matches 79; Conservative 38; Mismatches 132; Indels 121; Gaps 18;

QY 148 RRRPRRFVGENSDSEEDHPAFCDVPTQTGAESDSDGEGPSTRHSASG-----VQ 200  
Db 410 RTAPETVWGDGNSFDSFH-----LPTTRD--PGERCGRVTAADHAVGASLGVESLQ 462  
QY 201 P-----VDDANADSPGSDGEGPSTRHSASGPPADETTVHTDVEDD----- 242  
Db 463 PPTTLADERSGRSGDGRGSPSA-----ADDRVGVNNSDNPFTPTTPADDES 513  
QY 243 --LTLDKESACALMYHVGEMDMLRAMCDEDLFDLGIPEVIATSPQGGTDSAG-V 299  
Db 514 RRLSRLEGEQ-----DAGTASEDV--AVFGDSRRAGEK 546  
QY 300 VTEGSIASAASVGA-----GVEDVVLG-----ALEAQNVAAGEVYLEISDEEVD 343  
Db 547 VTHPTGSPARGSRSPGVPAED-----AGKASPGHRVPSARSQAQVSGFNQAEG 603  
QY 344 GAGLPP-----ASRRPRPVGS-----FLWDDGPRRH-ERTTTRIRHRKLSAYYRV 389  
Db 604 DEGLPPRDTDGRGAARRSPGGDGAFAFLNHGVDVNSHKEPPRPSE-----KT 654  
QY 390 ARPVMITDRLGVEVFYFGPAMSLVERKVFILCONPLADISHCLSRKGLRVLPLK 449  
Db 655 GEDPSMVKTAGDMSPLYGDPSTGLMHK-----CSDRPTFESIHACTRELDEMVL--- 706  
QY 450 PDNNNTGPGD 459  
Db 707 DEDGGSGCD 716

RESULT 5  
C71513  
hypothetical protein CT456 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C;Species: Chlamydia trachomatis  
C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C;Accession: C71513  
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
Science 282, 754-759, 1998  
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra  
A;Reference number: A71570; MUID:99000809; PMID:9784136  
A;Accession: C71513  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1005 <ARN>  
A;Cross-references: GB:AE001319; GB:AE001273; NID:G3328881; PIDN:AAC68056.1; PID:G332888  
A;Experimental source: serotype D, strain UW-3/Cx  
C;Genetics:  
A;Gene: CT456

Query Match 3.8%; Score 113.5; DB 2; Length 1005;  
Best Local Similarity 21.0%; Pred. No. 2.7;  
Matches 59; Conservative 41; Mismatches 88; Indels 89; Gaps 13;

QY 176 TQTGAESDSDG-----DEGPSTRHSASGVQ-PVDDANADSPGS-----GDEG----- 216  
Db 491 TQAGPSSDDGISFNETPCAGAAPAPSPPTPSIPILNVNVGVNIGNTVNTNT 550  
QY 217 -PSTRHSASGPPADETTVHTDVEDDLTLDKESACA-----LMYHVGQEMDMLM 266  
Db 551 TPTTQSTDASTDSTDDINTNNQTDINTDKSDGAGVNGDISSETSSGDDSGVS 610  
QY 267 RAMCDED-----LFDLLGTP-----DVATSPQGGTDA-----SGVTE 302  
Db 611 SSESDDKVASVNGDGPAMKOILSAVRKHLVDVYPGNGSGSTEGFLPANQTLGDIVSDENK 670  
QY 303 GS-----IAASAVGAGVEDVYLALAEQNVAGVYLEISDEEVDGGA 345  
Db 671 GSAQDTKLSNTGAGDDDDPTTAAVNGAEETLS-----DTDGIGDDVSDTSSGSDSG 727  
QY 346 GL---PPASRRPRPVGEFLWDDGP-----RRH 369

## RESULT 7

P84485  
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: P84485  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

Db 728 GVSSPSSSNKNTAVG-----NDGPSGLDILAAVRKH 759

## RESULT 6

A55620  
apical endosomal protein precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 24-Feb-1995 #sequence\_revision 24-Feb-1995 #text\_change 05-Nov-1999  
C;Accession: A55620  
R;Speelman, B.A.; Allen, K.; Grounds, T.L.; Neutra, M.R.; Kirchhausen, T.; Wilson, J.M.  
J. Biol. Chem. 270, 1583-1588, 1995  
A;Title: Molecular characterization of an apical early endosomal glycoprotein from devel  
A;Reference number: A55620; MUID:95130529; PMID:7823488  
A;Accession: A55620  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-1216 <SPE>  
A;Cross-references: GB:L37380; NID:9642652; PIDN:AAA65200.1; PID:G777776  
C;Superfamily: LDL receptor ligand-binding repeat homology; NAM homology  
F:231-267/Domain: LDL receptor ligand-binding repeat homology <LDLA>  
F:456-489/Domain: LDL receptor ligand-binding repeat homology <LDLB>  
F:654-813/Domain: NAM homology <MAM1>  
F:819-973/Domain: NAM homology <MAM2>

Query Match 3.7%; Score 111; DB 2; Length 1216;

Best Local Similarity 20.4%; Pred. No. 5.4;  
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;

QY 30 RSTGTFLLAPARNVDVPLDSLOFFIDFKRECLSKGLHPRDLLGSPITAFGKICTSTRRLR 89  
Db 708 RQGALLLTRQPVVP-----KECLSFYH---LHGPI---GTLCLAMR-- 748  
QY 90 RLPGEYEVV--QGNCRRRLILCAEVKECWVCWCHLHSGSSLWEIL-----YQ 139  
Db 749 --EGEEDTLNRSRGTHGRWH-----QAWTLH---HOLQSTKYQLLFEGRLRGYH 796  
QY 140 HSVRLKHKRRPRRPFVGENSDSEEDHPAFCDVPTQTGAESDSDGEGPSTR-HSASG 198  
Db 757 GTMGLLDMVAVRPGCMAAKRCSFEDSD---CGF-----SPDGWGLWTRQNNASG 842  
QY 199 VQP-----VDDANADSPGS---GDEGP-----STHSDSQPPADETTVHTDVEDDL 243  
Db 843 LGPWGFWIDHTTGTAGQHYVVDVTPNLLPKGVASLTSEEHP----- 886  
QY 244 TLDKESACALMYHVGQEMDMLRAMCDEDLFDLGIPEVDVIATSPQGGTDSAGVVT-- 301  
Db 887 --LSRPACLSFWYKLSFNPCTLVFVEE-----STRQELSSICHGGFAWRGLGSVNVQ 938  
QY 302 ---EGSIAASAVGAGVEDVYLALAEQNVAGVYLEISDEEVDGAGLPPAS----- 351  
Db 939 AEQAMKVVFEMASGVSESYNA-----LDDISLQDGPACQPGSCDFESGL 983  
QY 352 ---RRRPVVGEFLWD---DGPRRHERPTTRIRHRKLSAYYRVARPVMITDLGV 402  
Db 984 CGWSHLFPGLGGYSWWSGSGATSPRPFS-----VDHTVGT 1021  
QY 403 EVFYF-----GRPAMSLEVERKVFILCONPLADISHSL-----HSRKG 442  
Db 1022 EAGHFAPFETSVLPGCGGAALGSE-----PLPATAVSCLHFWYMGPPAFYKG 1071  
QY 443 -LRVLL 447  
Db 1072 ELRVLL 1077

## RESULT 7

P84485  
probable retroelement pol polyprotein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: P84485  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;





QY 227 PPA-----DETTVHTDNVEDDLTLLD-----KESACALMYH 257  
Db 188 PRAQFTSADAEEDDDDRDSEDETEAGDNGEDENGSEAGNRERRRRKAS-----242  
QY 258 VQEWDMMLMRANCDDELFDLLGIPEDVIATSPQGGDTAGGVVTEGSIASAVGAGVEDV 317  
Db 243 -GDNDAALEGFLDD-----PENTVHERVPRAGDKAGNSQDGG-----281  
QY 318 YLAGALAEONVAGEVYLEISDEEVDDGAGLPPASRRRPV---GEFLWDDGPRRH---ERP 372  
Db 282 --SSTIKIGDGTRELAQRDRD---DAQRPPVULSEAEFL---ARREAVRV 332  
QY 373 TTRIRHRKLSAYRVARPV-----MITDRIGVEVYFGRPAMSLVERKVFIIL 423  
Db 333 MVRDRVR-----TEPPLGTRVYQIAVLBD--GIVVEHFVTSAASLASVGNVYLG 381  
QY 424 CSQNPADISHSLHSRKLGLVLPKPDNNTPGVDVNLAAVLRSFASGLVIVLSRGI 483  
Db 382 IVQNVLPSEAFVDFIGRNGVL-----YAGEVNDVADGLGG-ADRKIEQALRPGD 432  
QY 484 YV 485  
Db 433 YV 434

RESULT 10  
B84417  
cell surface glycoprotein [imported] - Halobacterium sp. NRC-1  
C/Species: Halobacterium sp. NRC-1  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: B84417  
R/NG: W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L  
A/Title: Genome sequence of Halobacterium species NRC-1.  
A/Reference number: B84160; MUID:20504483; PMID:11016950  
A/Accession: B84417  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-836 <STO>  
A/Cross-references: GB:AE004437; NID:g10582055; PIDN:AAG20702.1; GSPDB:GN00138  
C/Genetics:  
A/Gene: csq  
C/Superfamily: S-layer glycoprotein

Query Match 3.6%; Score 107; DB 2; Length 836;  
Best Local Similarity 24.0%; Pred. No. 6.5;  
Matches 56; Conservative 30; Mismatches 67; Indels 80; Gaps 13;

QY 182 SEDSGDEGSPTRHSASGV-----QPVDNADSPSGDEG-----216  
Db 42 SEDGKTEG-----SVASGATIFQGBEDVTRFKLDNEKEVSPATLSRTGSGDEGVPLOMPIP 97  
QY 217 --PSTRHSDSOPPPADE---TTVHT-----DNVEDDLT-----LLDKESACAL-- 254  
Db 98 EDQSTGVSVDGPNDEADFGVTVQSPSVTLVVRNADNDVTGGVLTNQDESSIAVDY 157  
QY 255 MYHVQEMDMMLMRANCDDELFDLLGIPEDVIATSPQGG-----DTASG 298  
Db 158 NYIAAEDELATVE-----DEGLD---VTDEILAADQSGGAYEDGTGNGNGPNTLRFIDPNN 211  
QY 299 VYTEGSIASAVGAGVEDVYLAGALEONV---AGEVYLEISDEEVDDGAGL 347  
Db 212 -VDAGDYTVSV--EGVEDLDFGDATESASVTISSNKAISLAEDEVVQGANL 261

C/Accession: S57653  
R/Rauch, U.; Forsberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.  
submitted to the EMBL data Library, May 1995  
A/Description: Cloning and sequence of mouse neurocan and brevican and their different e  
A/Reference number: S57653  
A/Accession: S57653  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-883 <RAU>  
A/Cross-references: EMBL:X87096; NID:g886889; PIDN:CAA60575.1; PID:g886890  
C/Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG  
F/49-138/Domain: immunoglobulin homology <IMM>  
F/173-250/Domain: link protein repeat homology <LNK1>  
F/271-352/Domain: link protein repeat homology <LNK2>  
F/626-657/Domain: EGF homology <EGF>  
F/664-784/Domain: C-type lectin homology <LCH>  
F/791-847/Domain: complement factor H repeat homology <FHD>

Query Match 3.6%; Score 107; DB 2; Length 883;  
Best Local Similarity 24.0%; Pred. No. 7;  
Matches 78; Conservative 19; Mismatches 104; Indels 124; Gaps 16;

QY 151 PRPFVGEN-----SDSSEEDHFA-----PCDVPVTQTGAESDSDG-----DE 188  
Db 423 PRFLESETQSIAPPTSESEEGVALESEERFKDLEALEEKEQEDLWVWPRELSSPLPT 482  
QY 189 GPSTRHSASGVQVDDNADSPSGDEGSPTRHSQSPPPADETTVHTDNVEDDLTLLDK 248  
Db 483 GSETEHSLSQVSPPAQVQLDASPSGPPR---FRGPPAE-----TLLP- 524  
QY 249 ESACALMYHVQEMDMMLMRANCDDELFDLLGIPEDVIATSPQGGDTAGGVVTEGSIAS 308  
Db 525 -----PREWSATSTEGG-----AR 538  
QY 309 AVGAGVEDVYLAGALEONVAGEVYLEISDEEVDDGAGLPPASRRRPVVGEEFLWDD-GPR 367  
Db 539 EVGGTGSPELSPRESSEAGSSILE-----DGPSLLPAT-----WAPVGPR 581  
QY 368 RHERPTTRIRHRKLSAYRVARPVMITDRL---GVEVFYFGRPAMSLVERKVFIILC 424  
Db 582 ELETSPSEKSGRTVL--AGTSVQAQVPLPTDSASHGVAV-----APSSG-----DC 626  
QY 425 SQNPADISHSLHSRKLGLVLP 448  
Db 627 IPSPCNH-GGTCLKEKEGFRCLCLP 650

RESULT 12  
A40580  
lodestar maternal-effect protein - fruit fly (Drosophila melanogaster)  
N/Alternate names: probable nucleoside triphosphate binding protein lodestar  
C/Species: Drosophila melanogaster  
C/Date: 03-Aug-1992 #sequence\_revision 03-Aug-1992 #text\_change 21-Jul-2000  
C/Accession: A40580; S19008  
R/Girdham, C.H.; Glover, D.M.  
Genes Dev. 5, 1786-1799, 1991  
A/Title: Chromosome tangling and breakage at anaphase result from mutations in lodestar.  
A/Reference number: A40580; MUID:92009170; PMID:1916263  
A/Accession: A40580  
A/Molecule type: mRNA  
A/Residues: 1-974 <GIR>  
A/Cross-references: GB:X62629; NID:g8187; PIDN:CAA44496.1; PID:g8188  
C/Genetics:  
A/Gene: lodestar  
A/Cross-references: FlyBase:FBgn0002542

Query Match 3.6%; Score 107; DB 2; Length 974;  
Best Local Similarity 22.7%; Pred. No. 8;  
Matches 82; Conservative 52; Mismatches 149; Indels 78; Gaps 17;

QY 158 ENSDSSEED--HPAFCDVPVTQTG--AESEDSGDGE-----GPSTRHSASGVQPV---D 203  
Db 82 QSESEDSSEDSVSPSARNTKTKPLGTPSDESEDELEQALSPSTRMSITGVRPQDLSD 141



Search completed: June 9, 2004, 09:03:11  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:51:34 ; Search time 18 Seconds  
(without alignments)  
1640.210 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRLTWISFIVGALDSD.....QFDNVPVLVILRLRSVTCD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141691

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121.5	4.1	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
2	118	3.9	1004	1 PTPX_RAT	Q63475 rattus nov
3	116.5	3.9	466	1 XK10_MOUSE	Q04888 mus musculu
4	113.5	3.8	1005	1 Y456_CHLTR	Q84462 chlamydia t
5	113	3.8	1007	1 SAL2_HUMAN	Q9Y467 homo sapien
6	112.5	3.8	466	1 SAL2_HUMAN	Q9Y467 homo sapien
7	112.5	3.8	466	1 XK10_MOUSE	Q84462 chlamydia t
8	111	3.7	1161	1 NRDC_MOUSE	Q84462 chlamydia t
9	111	3.7	1216	1 AEGP_RAT	Q84462 chlamydia t
10	109.5	3.7	1004	1 SAL2_MOUSE	Q9Y467 homo sapien
11	108.5	3.6	503	1 IRT7_HUMAN	Q9Y467 homo sapien
12	107	3.6	883	1 PGB_MOUSE	Q9Y467 homo sapien
13	107	3.6	969	1 PTK_BIFLO	Q9Y467 homo sapien
14	107	3.6	1052	1 CLM_MOUSE	Q9Y467 homo sapien
15	107	3.6	1061	1 LDS_MOUSE	Q9Y467 homo sapien
16	106.5	3.6	704	1 CLIF_HUMAN	Q9Y467 homo sapien
17	106.5	3.6	1196	1 XPG_XENLA	Q9Y467 homo sapien
18	106.5	3.6	2333	1 PGCX_CANFA	Q9Y467 homo sapien
19	105.5	3.5	738	1 S521_RAT	Q9Y467 homo sapien
20	104.5	3.5	1161	1 NRDC_RAT	Q9Y467 homo sapien
21	104.5	3.5	5596	1 MDN1_HUMAN	Q9Y467 homo sapien
22	104	3.5	852	1 CSG_HALHA	Q9Y467 homo sapien
23	103.5	3.5	273	1 IE68_HSVB4	Q9Y467 homo sapien
24	103	3.4	950	1 URBI_USPTA	Q9Y467 homo sapien
25	102.5	3.4	2124	1 PGCX_RAT	Q9Y467 homo sapien
26	102	3.4	838	1 TAC3_HUMAN	Q9Y467 homo sapien
27	102	3.4	1085	1 IFH1_YEAST	Q9Y467 homo sapien
28	101.5	3.4	494	1 SOX9_CHICK	Q9Y467 homo sapien
29	101.5	3.4	562	1 STL1_HUMAN	Q9Y467 homo sapien
30	101.5	3.4	1239	1 M4K4_HUMAN	Q9Y467 homo sapien
31	101	3.4	3415	1 POLG_POWVL	Q9Y467 homo sapien
32	100.5	3.4	1966	1 CCAP_HUMAN	Q9Y467 homo sapien
33	100.5	3.4	2459	1 MAPB_RAT	Q9Y467 homo sapien

34	99.5	3.3	426	1 ICSB_HUMAN	Q02556 homo sapien
35	99.5	3.3	1601	1 PGR2_HUMAN	Q84847 homo sapien
36	99.5	3.3	2364	1 PGCA_BOVIN	Q13608 bos taurus
37	99	3.3	1058	1 BUB1_MOUSE	Q08901 mus musculu
38	98.5	3.3	456	1 EYZ2_CHLRE	Q08355 chlamydomon
39	98.5	3.3	695	1 ARP_EUGGR	Q04732 euglena gra
40	98	3.3	461	1 SX10_CHICK	Q9W757 gallus gall
41	98	3.3	2468	1 MAPB_HUMAN	Q9W757 gallus gall
42	98	3.3	4625	1 DY1A_CHLRE	Q9W757 gallus gall
43	97.5	3.3	536	1 CH62_CHLVC	Q9W757 gallus gall
44	97.5	3.3	1131	1 MYPF_CHICK	P58698 chlamydomon
45	97	3.2	528	1 RPSA_MYCTU	P16419 gallus gall
					Q60162 mycobacteri

#### ALIGNMENTS

RESULT 1  
NCR2\_HUMAN  
ID NCR2\_HUMAN STANDARD; PRT; 2517 AA.  
AC Q9Y618; Q00613; Q15416; Q13354; Q9Y5U0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nuclear receptor co-repressor 2 (N-COR2) (Silencing mediator of  
DE retinoic acid and thyroid hormone receptor) (SMRT) (SMRTE) (Thyroid-,  
DE retinoic-acid-receptor-associated co-repressor) (T3 receptor-  
DE associating factor) (TRAC) (CTG repeat protein 26).  
GN NCR2 OR CTG26.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
RC TISSUE=Pituitary;  
RX MEDLINE=99178941; PubMed=10077563;  
RA Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;  
RL "Unique forms of human and mouse nuclear receptor corepressor SMRT";  
Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
RC TISSUE=Cervical adenocarcinoma;  
RX MEDLINE=99199215; PubMed=10097068;  
RA Park E.J., Schreen D.J., Yang M., Li H., Li L., Chen J.D.;  
RL "SMRte, a silencing mediator for retinoid and thyroid hormone  
RT receptors-extended isoform that is more related to the nuclear  
RT receptor corepressor.";  
Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).  
RN [3]  
RP SEQUENCE OF 1023-2517 FROM N.A.  
RC TISSUE=Cervical adenocarcinoma;  
RX MEDLINE=96008552; PubMed=7566127;  
RA Chen J.D., Evans R.M.;  
RL "A transcriptional co-repressor that interacts with nuclear hormone  
RT receptors.";  
Nature 377:454-457(1995).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM TPAC-1).  
RC TISSUE=Fetal liver;  
RX MEDLINE=96408715; PubMed=8813722;  
RA Sande S., Privalsky M.L.;  
RL "Identification of TRACs (T3 receptor-associating cofactors), a family  
RT of cofactors that associate with, and modulate the activity of,  
RT nuclear hormone receptors.";  
Mol. Endocrinol. 10:813-825(1996).  
RN [5]  
RP SEQUENCE OF 428-613 FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=97369492; PubMed=9225980;  
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,  
RA Breschel T.S., Steine O.C., Callahan C., McInnis M.G., Ross C.A.;  
RL "CDNAs with long CAG trinucleotide repeats from human brain.";

RL Hum. Genet. 100:114-122(1997).  
 RP [6]  
 RP INTERACTION WITH MINT.  
 RX MEDLINE=21231190; PubMed=11331509;  
 RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,  
 RA Hon M., Evans R.M.;  
 RA "Sharp, an inducible cofactor that integrates nuclear receptor  
 RT repression and activation.";  
 RL Genes Dev. 15:1140-1151(2001).  
 RN [7]  
 RP INTERACTION WITH HDAC10.  
 RP MEDLINE=21839031; PubMed=11739383;  
 RA Fischer D.D., Cai R., Bhatia U., Asselbergs F.A.M., Song C., Terry R.,  
 RA Trognani N., Widmer R., Adadja P., Cohen D.;  
 RT "Isolation and characterization of a novel class II histone  
 RT deacetylase, HDAC10.";  
 RL J. Biol. Chem. 277:6656-6666(2002).  
 CC -1- FUNCTION: Mediates the transcriptional repression activity of some  
 CC nuclear receptors by promoting chromatin condensation, thus  
 CC preventing access of the basal transcription.  
 CC -1- SUBUNIT: Interacts with HDAC7 (By similarity). Forms a large  
 CC corepressor complex that contains SIN3A/B and histone deacetylases  
 CC HDAC1 and HDAC2. This complex associates with the thyroid (TR) and  
 CC the retinoid acid receptors (RAR) in the absence of ligand, and  
 CC may stabilize their interaction with TFIIB. The SRMT isoform  
 CC interacts with HDAC10. Interacts with MINT.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Name=Alternative splicing; Named isoforms=2;  
 CC Name=SMRT; Synonyms=TRAC-2;  
 CC IsoId=Q9Y618-1; Sequences=Displayed;  
 CC Names=TRAC-1;  
 CC IsoId=Q9Y618-2; Sequences=VSP\_003412, VSP\_003413;  
 CC Note=Contains only the C-terminal receptor-interacting domain  
 CC and acts as an antirepressor;  
 CC -1- TISSUE SPECIFICITY: Ubiquitous. High levels of expression are  
 CC detected in lung, spleen and brain.  
 CC -1- INDUCTION: Regulated during cell cycle progression.  
 CC -1- DOMAIN: The N-terminal region contains repression functions that  
 CC are divided into three independent repression domains (RD1, RD2  
 CC and RD3). The C-terminal region contains the nuclear receptor-  
 CC interacting domains that are divided in two separate interaction  
 CC domains (ID1 and ID2).  
 CC -1- DOMAIN: The two interaction domains (ID) contain a conserved  
 CC sequence referred to as the CORNR box. This motif is required and  
 CC sufficient to permit binding to unliganded TR and RARs. Sequences  
 CC flanking the CORNR box determine nuclear hormone receptor  
 CC specificity.  
 CC -1- SIMILARITY: Contains 1 SANT-A domain.  
 CC -1- SIMILARITY: Contains 1 Myb-like domain.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF113003; AAC20946.1; -;  
 DR EMBL; AF125672; AAC22973.1; -;  
 DR EMBL; U37146; AAC50236.1; -;  
 DR EMBL; S83390; AAB50847.1; -;  
 DR EMBL; U80750; AAB91446.1; -;  
 DR PIR; S60255; S60255.  
 DR PDB; 1KKQ; 27-FEB-02.  
 DR TRNSPAC; T04689;  
 DR Genew; HGNC:7673; NCOR2.  
 DR MIM; 600848; -;  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0003714; F:transcription co-repressor activity; TAS.

DR InterPro; IPR001005; Myb DNA binding.  
 DR Pfam; PF00249; myb DNA-binding; 2.  
 DR SMART; SM00717; SANT; 2.  
 DR PROSITE; PS0090; MYB\_3; 1.  
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 KW Coiled coil; Alternative splicing; 3D-structure.  
 FT DOMAIN 174 215  
 FT DOMAIN 254 312  
 FT DOMAIN 429 474  
 FT DNA\_BIND 613 657  
 FT DNA\_BIND 522 561  
 FT DOMAIN 778 820  
 FT DOMAIN 2139 2143  
 FT DOMAIN 2342 2346  
 FT DOMAIN 494 510  
 FT DOMAIN 682 685  
 FT DOMAIN 994 1002  
 FT DOMAIN 1384 1389  
 FT DOMAIN 1842 1846  
 FT DOMAIN 2479 2482  
 FT VARSPLIC 1 1702  
 FT VARSPLIC 2353 2398  
 FT CONFLICT 7 7  
 FT CONFLICT 295 295  
 FT CONFLICT 309 309  
 FT CONFLICT 352 352  
 FT CONFLICT 365 365  
 FT CONFLICT 612 613  
 FT CONFLICT 711 711  
 FT CONFLICT 724 740  
 FT CONFLICT 787 796  
 FT CONFLICT 804 804  
 FT CONFLICT 814 814  
 FT CONFLICT 817 817  
 FT CONFLICT 889 889  
 FT CONFLICT 1023 1030  
 FT CONFLICT 1034 1034  
 FT CONFLICT 1894 1894  
 FT CONFLICT 2494 2494  
 SQ SEQUENCE 2517 AA; 274031 MW; F5905C01761258C0 CRC64;  
 Query Match 4.1%; Score 121.5; DB 1; Length 2517;  
 Best Local Similarity 21.3%; Pred No 2.4;  
 Matches 8; Conservative 48; Mismatches 132; Indels 119; Gaps 17;  
 Qy 105 RWRLLCAEV--KECWCVHARTHLHSGSLWILYQHSVRLKRRRRPRPFVGENSDS 162  
 Db 632 RWSAIAARMVGSKTQSKNFYENYKQNLDEILOQHLKMKERNARKKKAPAAAS 691  
 Qy 163 SEEDHPAFCDVPVTQ-----TCAESE-----DSGDE-----GPSTRHAS 197  
 Db 692 EEAAP-----PVDEEMASGVSGNEEMVEAEALHASGNEVPRGCSGPATVNNSS 746  
 Qy 198 GVQ-----PVDDANADFGSGDEGFSTRHSDSOPP-----PADETT--- 233  
 Db 747 DTESISPHTEAAKDGQNGKPPATLGADGPPGPPPTPRRTSRAPISPTPASEATGAP 806  
 Qy 234 -----VHTDNVEDDLTLDKESACALMYHVQEMDMMLRAMCDEDLFDLL 278  
 Db 807 TPPPAPSPSAPPVVPVKEKEE-----ETAAAPFVEEGE-----QKPPAAEELAVDT 855  
 Qy 279 GIPEDVIAT-----QPGGDTDASGVVTEGSTAASAVGAGVEDVYLAGALEAQNVA 329  
 Db 856 GKAEFPVKSECTEAEEGPAKGDAAEATAEGALKAEKKEGG-----SGRAATTAKS 909  
 Qy 330 G-----EYVLEISDENVDGAG-----LPPASRRRPVVGEFLWDDGPPRRHERP-TTRI 377  
 Db 910 GAPQDSSTATCSADEVDEAEAGDKNRLSPRPSLITPTGD-----PRANASPOKPLDL 963  
 Qy 378 RHRKLRSAYRVARPPVMLT 397

Db 964 KOLKORAA----AIPPIQVT 979

RESULT 2

FTPX\_RAT STANDARD; PRT; 1004 AA.

AC Q63475;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Receptor-type protein-tyrosine phosphatase N2 precursor (EC 3.1.3.48)

DE (R-PTP-N2) (PTP NE-6) (PTPNE6) (Phogrin)

GN PTPN2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=New England Deaconess Hospital; TISSUE=Insulinoma;

RA Wasmeyer C., Hutton J.C.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley;

RA Fitzgerald L.R., Walton K.M., Dixon J.E., Largent B.L.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Implicated in development of nervous system and pancreatic endocrine cells.

CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein tyrosine + phosphate.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).

CC -1- TISSUE SPECIFICITY: Pancreas.

CC -1- PTM: Appears to undergo multiple proteolytic cleavage at consecutive basic residues (By similarity).

CC -1- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z50735; CAA90600.1; -

DR EMBL; U73458; AAC08036.1; -

DR HSP; P18052; IYFO.

DR InterPro; IPR000387; TYR phosphatase.

DR InterPro; IPR000242; Tyr\_PP.

DR Pfam; PF00102; Y\_phosphatase; 1.

DR PRINTS; PR00700; PRTYPHPTASE.

DR SMART; SM00194; PTPc; 1.

DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.

DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.

DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.

KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane.

FT SIGNAL 1

FT CHAIN 28 1004 RECEPTOR-TYPE PROTEIN-TYROSINE PHOSPHATASE N2.

FT DOMAIN 28 603 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 604 624 POTENTIAL.

FT DOMAIN 625 1004 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 759 993 PROTEIN-TYROSINE PHOSPHATASE.

FT ACT\_SITE 934 934 PHOSPHOCYSTEINE INTERMEDIATE (POTENTIAL).

FT SITE 415 416 CLEAVAGE SITE (BY SIMILARITY).

FT CARBOHYD 553 553 N-LINKED (GLCNAC...) (POTENTIAL).

SQ SEQUENCE 1004 AA; 111863 MW; A73929E11B486FB2 CRC64;

Query Match 3.9%; Score 118; DB 1; Length 1004;

Best Local Similarity 21.8%; Pred.No. 1.3; Indels 198; Gaps 27;

Matches 115; Conservative 60; Mismatches 159;

QY 153 RPFVGENSDSEEDHPAFCDYPTVOTGAESDESDGDEGSTR-----HSASGVQPVYD- 203

Db 177 RPKVGE--DSSPENILTY-----VAHTSALTVP-----PATRVKYPDNLLRPLSLQPDDEL 225

QY 204 ----DANADS-----PGSGDEGSPSTRHSDSQP----- 226

Db 226 SPKVDSDIDKQKLI AALGAYTAQRPPGEND--PEPRYL VHSMPRAPFAAPALSQRWPL 283

QY 227 PADETTVHTDNVEDDLILLKESACALMYHVGQ--EMDMLMRAMCDEDLFDLLG-IPED 283

Db 284 PPGDS-----KDSLSMGDDTLRLSLKLOLQQAQAEVDRLGSLKLEEQADSIAQAIQSD 335

QY 284 VIATSPQGGDTDASGVTWTEGSIASAVGAGVEDVYLAGALEAQNAGVYVLEISDE--E 340

Db 336 PVEGSO-----ZSHRGAE-----CQLREQADAPEMLQ--DRLRPE 370

QY 341 VDDGAGLPPASRRRPPVVGFEFLWDDGP-----RRHERP-----TTR 375

Db 371 VDDPAAYKEVSRSLSPKGLDGLLDHSGPLLPAPLEKSSRAEMKKSEQPEEVLSEEBETA 430

QY 376 RIRHREKLS-----AYRYA-----RPPVWITDRLGVEVYFGRPA--MSL 414

Db 431 GVEHVKSRTYSKDLLERKPNSEPOFWRLDFOFQRAPEVWEDEQNKLKLAAGQPPSGGQL 490

QY 415 EV-----ERKVFILCSQNPLA-----DISHSLHSRKLRLVLPKP-----DDNNT 455

Db 491 EVQPEEEOQGVILGNPNLSPKGLQMLDEVAH-----LLRVPSSFFADVKVL 539

QY 456 GPGDVNLLAALVRSF-----ASGLVIVLSRSGIYVYVKNLCKSVLVHVG 497

Db 540 GPAVIFKVSANIQNTTADVTKAAVDNKKDELEKATGLTI--LQSGIRPKGKLKLLPHPEE 597

QY 498 NNPPKKFGVICGLSSRAVLDFVNAQ-----YRIQHEHIKKTTFVIGGDPTS 545

Db 598 QEDSKFVILFSLACILAVLLASSLAYCLRHNSHYKLEKLSGLGADPSA 649

RESULT 3

SX10\_MOUSE STANDARD; PRT; 466 AA.

ID SX10\_MOUSE

AC Q04888; O08518; O09141; O54856; P70416;

DT 01-JUN-1994 (Rel. 29, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Transcription factor SOX-10 (SOX-21) (transcription factor SOX-M).

GN SOX10 OR SOX-10 OR SOX21.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RC MEDLINE=98085871; PubMed=9425902;

RA Southard-Smith E.M., Kos L., Pavan W.J.;

FT "Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model.";

RL Nat. Genet. 18:60-64 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RC MEDLINE=98226785; PubMed=9560246;

RA Herbarth B., Pingault V., Bondurant N., Kuhlbrodt K.,

RA Hermans-Borgmeyer I., Puliti A., Lemort N., Goossens M., Wegner M.;

RT "Mutation of the hry-related Sox10 gene in dominant megacolon, a mouse model for human Hirschsprung disease.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:5161-5165 (1998).

RN [3]

RP SEQUENCE FROM N.A.

RC MEDLINE=98430992; PubMed=9760192;

RA Pusch C., Huestert E., Pfeifer D., Sudbeck P., Kist R., Roe B.,

RA Wang Z., Balling R., Blin N., Scherer G.;

RT "The SOX10/Sox10 gene from human and mouse: sequence, expression, and

transactivation by the encoded HMG domain transcription factor.";  
Hum. Genet. 103:115-123(1998).

[4]

SEQUENCE FROM N.A.

STRAIN=C3H/HeN;

MEDLINE=97179197; PubMed=9027483;

Tani M., Shindo-Okada N., Hashimoto Y., Shiroishi T.,

Takenoshita S., Nagamachi Y., Yokota J.,

"Isolation of a novel Sry-related gene that is expressed in high-

metastatic K-1735 murine melanoma cells.";

Genomics 39:30-37(1997).

[5]

SEQUENCE FROM N.A.

MEDLINE=22388257; PubMed=12477932;

Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,

Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,

Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

Raha S.S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

Fahy J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,

Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,

Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,

"Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[6]

SEQUENCE OF 112-167 FROM N.A.

MEDLINE=93181275; PubMed=8441686;

Wright E.M., Snopce B., Koopman P.,

"Seven new members of the Sox gene family expressed during mouse

development.";

Nucleic Acids Res. 21:744-744(1993).

[7]

SEQUENCE OF 114-167 FROM N.A.

STRAIN=Swiss Webster; TISSUE=Embryonic tooth;

MEDLINE=97079683; PubMed=8921394;

Stock D.W., Buchanan A.V., Zhao Z., Weiss K.M.,

"Numerous members of the Sox family of HMG box-containing genes are

expressed in developing mouse teeth.";

Genomics 37:234-237(1996).

FUNCTION: Transcription factor that seems to function

synergistically with the POU domain protein TST-1/OCT6/SCIP. Could

confer cell specificity to the function of other transcription

factors in developing and mature glia (By similarity).

SUBCELLULAR LOCATION: Nuclear.

DISEASE: Defects in SOX10 are the cause of the mouse mutant

dominant megacolon (dom). While dom/+ heterozygous mice display

regional deficiencies of neural crest-derived enteric ganglia in

the distal colon, dom/dom homozygous animals are embryonic lethal.

SIMILARITY: Contains 1 HMG box domain.

CAUTION: Ref 4 sequence differs from that shown due to

frameshifts in positions 16 and 392.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; AF017182; AAB95738.1; ALT\_INIT.

EMBL; AF047043; AAC24564.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

EMBL; U66141; AAB49282.1; ALT\_FRAME.

DR EMBL; BC018551; AAB18551.1; --  
DR EMBL; Z18959; CAA79484.1; --  
DR EMBL; U70441; AAC52859.1; --  
DR PIR; S30242; S30242.  
DR HSP; P48436; ISX9.  
DR MGD; MG198358; Sox10.  
DR GO; GO:0030154; P:cell differentiation; IMP.  
DR InterPro; IPR000910; HMG box; 1.  
DR Pfam; PF00505; HMG box; 1.  
DR SMART; SM00398; HMG; 1.  
DR PROSITE; PS01118; HMG\_BOX\_2; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein.  
FT DOMAIN 35 41  
FT DNA BIND 104 172  
FT VARIANT 11 11 E -> V.  
FT CONFLICT 41 41 G -> V (IN REF. 4).  
FT CONFLICT 119 119 R -> P (IN REF. 6).  
FT CONFLICT 357 357 K -> E (IN REF. 4).  
SQ SEQUENCE 466 AA; 49949 MW; 84862547F039E71F CRC64;  
  
Query Match 3.9%; Score 116.5; DB 1; Length 466;  
Best Local Similarity 19.7%; Pred. No. 0.58;  
Matches 81; Conservative 46; Mismatches 147; Indels 137; Gaps 19;  
  
QY 119 WCYHAR-----THLHSG-----SSLWEILYQHSVRLEKRRPRPRPFVGENSP---S 162  
Db 114 WQAARRLADQYPHLHNAELSKTLGLKRLNESD-----KRFIEAEERLRVQ 163  
QY 163 SEDHPAFQCPVPTQ-----TGAESEDSGDEGPST-----RHSASGVQP 201  
Db 164 HKKDHPDYQYPRRRKNGKAAQGEAEPCGAEAGGAAAIQAHYKSAHLDRHPEEG-SP 222  
QY 202 VDDANADSPGSGDEGSPSTRHSQPPADETV-----HTNVED 241  
Db 223 MSDGNPEHSPGSHGPPT-----PPTTKTELQSGKADPKDGRSLGEGKPHIDFGNV 276  
QY 242 DLTLLDKESACALMYHVQGMMLKRAMCDEDLFLLGIPEDVIATSQGGTDSAGVVT 301  
Db 277 DGEISHE-----VMSNMETFDVTELDQYLPNGHFG-----H 309  
QY 302 EGSIAASAVGAGVEDVYLAGALEAQNAGEYVLEISDEVDGAGLPPAS-----RRRPV 356  
Db 310 VGSYSAAGYGLG-----SALAVASGHSAWISK-----PPGVALPTVSPGVDAAQV 356  
QY 357 VGEFLWDDGPRRH-ERPTTRRIRHRKLR-----SAYYVARPPVMTDLRGVEVFGRP 410  
Db 357 KTTCTPQGPPTDQPSQIATYSLSLPHYGSAPPSISRQFDYSDHQPSGP-YIGHA 415  
QY 411 AMSLEY-----ERKVFILCSQ-NPLADISHSLSRKGLRVLKPK 450  
Db 416 GQASGLYSAFSYMGSPORPLYTASDPSPSGQSHSPHWEQPVVTLTSLRP 466  
  
RESULT 4  
Y456\_CHLTR STANDARD; PRT; 1005 AA.  
AC 084462;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protein CT456 precursor.  
GN CT456.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=D/UV-3/CX;  
RX MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:





```

FT CONFLICT 554 554 V -> M (IN REF. 1).
FT CONFLICT 575 581 FRYLEP -> LPLCARA (IN REF. 1).
SQ SEQUENCE 1007 AA; 105398 MW; 24B7C4ADE852B96C CRC64;

Query Match
Best Local Similarity 19.4%; Score 113; DB 1; Length 1007;
Matches 111; Conservative 61; Mismatches 171; Indels 228; Gaps 25;

QY 146 KRRRPRR-----PFVGENSDSEEDHPACDFVVTQ-----TGAESD----- 184
DQ 4 RKQRQQIISCEGSPASENGDAEEDHPVCARCCQAQFTDTEFLAHQACSTDPVPM 63
QY 185 ---SGDEGSTRHSAGVOP-----VDDANADSPSGDE-----GPSTRHSQD- 225
DQ 64 VIIGQENFNSSASSEPPGEGHNPQVMDTSHNPPDSSGVPTDPTWGERRGESSG 123
QY 226 -----PPADETTVHTDNVE----- 240
DQ 124 HFLVAATGTAAGGGGLILASPGLGATPLPPESTPAPPPPPPPPPGVSGLNLIPLIL 183
QY 241 DDLTLLDK-----ESAC---ALMYHVGO-----EMDMLMEAMCDEDLFDLLG 279
DQ 184 EELRVLQQRQIHMOMQTEICQVLLGSLGQTVGAPASPSLPTGTASSTKPLPLFS 243
QY 280 IPEDV-----IATSPQSGDT-----DASGV----- 299
DQ 244 PIKPVQTKTLASSSSSSSSSSGAETPKQAFPHLYPLGSHQHPFSAGGVGRSHKPTAPS 303
QY 300 -----VTGSTRASAVGA---GVEDVYLAGALEAONVAGEYVLEISD 338
DQ 304 PALPGSTDQLIASPHLAFPTTGLLAAQCLGAARLEATASPLGLPKRNGSG-----ELSY 359
QY 339 EYVDDGAGLPPASRRPVVGEFLWDGP-----RHH---ERP-----TTR---RIR 378
DQ 360 GEWVGLEKPGGRHKRCFCFAKVGSDSALQIHLRSHTGERPYKCVGNFRFTRGNLKVH 419
QY 379 HKLRSAIYRVARPPVMTDRIGVEYFVYGRP-ANSL-----VERKVFIL 423
DQ 420 FRRHREKYPHVQNPVHPVDEHLDYVITSSGLPYGMSVPEKAEBAEATPGGQVERKPLVA 479
QY 424 CSQNPADISHSLCHSRKG-----LRVLLPK-PDNNNGTGP-DVNLLAAV 466
DQ 480 SITALSATSLTLTSLTASATATAPGLPAPNKVFLMKVAPKADENFPFSGSAGSV 539
QY 467 LRSFAS-----GLVIVLSRSGIYVKNLCKST 492
DQ 540 ABSSTATRMQLSKLVTSLPSWALLTNHFKST 570

RESULT 6
SX10_HUMAN STANDARD; PRT; 466 AA.
AC P56693;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-10.
GN SOX10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT WS4 LEU-ARG-161 INS.
RX MEDLINE=98122581; PubMed=9462749;
RA Pingault V., Bondurand N., Kuhlbrodt K., Goerich D.E., Prehu M.O.,
RA Puliti A., Herbarth B., Hermans-Borgmeyer I., Legius E., Matthijs G.,
RA Amiel J., Lyonnet S., Ceccherini I., Romeo G., Clayton Smith J.,
RA Read A.P., Wegner M., Goossens M.;
RT "SOX10 mutations in patients with Waardenburg-hirschsprung disease.";
RL Nat. Genet. 18:171-173(1998).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=98430992; PubMed=9760192;
RA Pusch C., Rustert E., Pfeifer D., Sudbeck P., Kist R., Roe B.,
RA Wang Z., Bailing R., Blin N., Scherf G.;
RT "The SOX10/Sox10 gene from human and mouse: sequence, expression, and
transactivation by the encoded HMG domain transcription factor.";
RL Hum. Genet. 103:115-123(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgen A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kersey J., Kimberley A.W., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay Y., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon I., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuayama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Hinds K., Kemp K., Lareille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker J.A., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Kori I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edlmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta, and Skin;
MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Besak A.S., Mcswain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

```

RT human and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences".  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

[5]

RP CHARACTERIZATION.

RX MEDLINE=98389729; PubMed=9722528;

RA Kuhlbrodt K., Schmidt C., Sock E., Pingault V., Bondurand N.,

RA Goossens M., Wegner M.;

RT "Functional analysis of Sox10 mutations found in human Waardenburg-

RT Hirschsprung patients".;

RL J. Biol. Chem. 273:23033-23038 (1998).

[6]

RN NUCLEOCYTOPLASMIC SHUTTLEING.

RP MEDLINE=22133596; PubMed=12138193;

RA Rehberg S., Lischka P., Glaser G., Stamminger T., Wegner M.,

RA Rosorius O.;

RT "Sox10 is an active nucleocytoplasmic shuttle protein, and shuttling

RT is crucial for Sox10-mediated transactivation".;

RL Mol. Cell. Biol. 22:5826-5834 (2002).

CC -!- FUNCTION: Transcription factor that seems to function

CC synergistically with the POU domain protein TST-1/OCT6/SCIP. Could

CC confer cell specificity to the function of other transcription

CC factors in developing and mature glia (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.

CC -!- TISSUE SPECIFICITY: Expressed in fetal brain and in adult brain,

CC heart, small intestine and colon.

CC -!- DISEASE: Defects in SOX10 are a cause of Waardenburg syndrome

CC type IV (WS4) [MIM:277580]; also known as Waardenburg-Shah

CC syndrome. WS4 is characterized by the association of Waardenburg

CC features (depigmentation and deafness) and the absence of enteric

CC ganglia in the distal part of the intestine (Hirschsprung

CC disease).

CC -!- SIMILARITY: Contains 1 HMG box domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; AJ001183; CRA04576.1; -

CC EMBL; AL031587; CAB62982.1; -

CC EMBL; BC002824; AA072824.1; -

CC EMBL; BC007595; AA07595.1; -

CC HSP; P48436; ISX9.

CC TRANSFAC; T04918; -

CC Genew; HGNC:11190; SOX10.

CC MIM; 602229; -

CC MIM; 277580; -

CC GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.

CC GO; GO:0003713; F:transcription co-activator activity; TAS.

CC GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.

CC GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.

CC InterPro; IPR000910; HMG\_12\_box.

CC Pfam; PF00505; HMG\_box; 1.

CC SMART; SMO0398; HMG; 1.

CC PROSITE; PS00118; HMG\_BOX\_2; 1.

CC Transcription regulation; DNA-binding; Nuclear protein;

CC Disease mutation; Hirschsprung disease; Deafness.

CC DOMAIN 134 145 NUCLEAR EXPORT SIGNAL.

CC FT DOMAIN 35 41 POLY-GLY.

CC FT DNA\_BIND 104 172 HMG BOX.

CC FT VARIANT 161 161 R -> RLR (in WS4).

CC FT FTID=VAR\_003743.

CC SEQUENCE 466 AA; 49911 MW; FAA1EC108D4DE6A1 CRC64;

Query Match 3.8%; Score 112.5; DB 1; Length 466;

Best Local Similarity 19.58; Pred.No.1.1;

Matches 80; Conservative 46; Mismatches 148; Indels 137; Gaps 19;

QY 119 WCWVAR-----THLHSG-----SSLWEILYQHSVRLEKRRRRPRFPVGENSD---S 162

Db 114 WAAARRLKLAQYPHLHNAELSKTLGKLRLLNESD-----KRPFFIEAEARLRMQ 163  
 QY 163 SEEDHPAFCDVPTQ-----TCABSESDSGDEGPST-----RHSAAGVQP 201  
 Db 164 HKKDHDPYKYQRRRKNGKAAQGEAEQGGTAIAQHYKSAHLDRHPGEG-SP 222  
 QY 202 VDDANADSPGSDGEGPSTRHSDSQPPPADETVV-----HTDNVED 241  
 Db 223 MSDGNPEHPSGQSHGPPT-----PPTTKTELOSGKADPKRDKRSMGEGKPHIDFGNV 276  
 QY 242 DLTLLDKESACALMYHVGQEMLMRAMCDLFDLLGIPEDVIATSPGGDTDASGVVT 301  
 Db 277 DIGEISHE-----VMSNMETFDVAELDQYLPNGHPG-----H 309  
 QY 302 EGSIAASAVGAGVEDVYLGALEAQNAGEVYELISDEVEDDGAGLPPAS-----RRRPV 356  
 Db 310 VSSYGAAGYGLG-----SALAVASGHSAWISK---PPGVALPTVSPGVDAKAQV 356  
 QY 357 VGEFLWDGDPRRH-ERPTTRIRHKLK-----SAYYHVARPPVMTDLGVENVYFGRP 410  
 Db 357 KTETAGPQGPQPHYDTPSTSQIATYSLPHYGAFFSISRPFQDYSRHQSPGP-YIGH 415  
 QY 411 AMSLEV-----ERKVFILCSQ-NPLADISHSLHRSKRLVLLPKP 450  
 Db 416 GOASGLYSAFSGVMSQRPQPLVTAISDPSPSGPQSHSPHWEQPVVYTLRSR 466

# RESULT 7

SOX10\_RAT  
 ID SX10\_RAT STANDARD; PRT; 466 AA.  
 AC O55170;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Transcription factor SOX-10.  
 GN SOX10.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=98075122; PubMed=9412504;  
 RA Kuhlbrodt K., Herbarth B., Sock E., Hermans-Borgmeyer I., Wegner M.;  
 RT "Sox10, a novel transcriptional modulator in glial cells.";  
 RL J. Neurosci. 18:237-250 (1998)  
 CC -!- FUNCTION: Transcription factor that seems to function  
 CC synergistically with the POU domain protein TST-1/OCT6/SCIP. Could  
 CC confer cell specificity to the function of other transcription  
 CC factors in developing and mature glia.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Predominant expression in glial cells of the  
 CC nervous system.  
 CC -!- SIMILARITY: Contains 1 HMG box domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

35

41

FT

FT DNA BIND 104 172 HMG BOX.  
SQ SEQUENCE 466 AA; 50039 MW; 094B0C5E42D83454 CRC64;  
Query Match  
Best Local Similarity 3.8%; Score 112.5; DB 1; Length 466;  
Matches 83; Conservative 50; Mismatches 143; Indels 133; Gaps 21;  
QY 119 WCVHAR-----THLHSG-----SSLWEILYQHSVLEKRRRPRPPFVGENSD--S 162  
DB 114 WAQAARRKLADQYPLHNAELSKTLGLWLLNESD-----KEPFTIEAERLRMQ 163  
QY 163 SEEDHPACDVVTQTG-----AESDSG-----DEGPST-----RHSASGVQP 201  
DB 164 HKXDHDPYKYPRRRKNGKAAQACFCGETDQGAAGAAIQAHYKSAHLDRHPPEEG-SP 222  
QY 202 VDNANADSPGSGDEGFSRHSQPPADETTV-----HTDNVED 241  
DB 223 MSDGNPEHPSGSHGPP-----PPTPKELQSGKADPKDRSGLSGEGKPHIDPQNV 276  
QY 242 DUTLKDSEACALMYHVGQEMDMLRMCDEDLFDLLGIPREDVIATSPG--GPTDASGV 299  
DB 277 DIGEISHE-----VMSNMETFDVTELDQYLPNGHPGHVGSYSAAGY 318  
QY 300 VTEGSIASAVGAGVEDVYLAGALEAQNVAQVYVLEISDEVDGAGLPPA-SRRRPVWG 358  
DB 319 ---GLSSALAVSG-----HSWISKPPGVALPTV-----SPPAVDAAQVKT 358  
QY 359 EFLWDGPRRH-ERPTTRIRRRKLRL-----SAYRVARPPVPMITDRLGVEVFGRPAM 412  
DB 359 ETTGQGGPHYTDQSTQSLAYTSLSLPHYSARFSPISRPQFSDHQSPG-YYGHAGQ 417  
QY 413 SLEV-----ERKVFILCSQ-NPLADTSHSLHRSKGLRVLLPKP 450  
DB 418 ASGLYSAFSYMGPSORPLTAISDPSGPOSHSPHWEQVYVYTLGRP 466

## RESULT 8

ID NRDC MOUSE STANDARD; PRT; 1161 AA.  
AC Q8BGL1;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nardilysin precursor (EC 3.4.24.61) (N-arginine dibasic convertase)  
DE (NRD convertase) (NRD-C).  
GN NRD1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=22354683; PubMed=1246851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,  
RA Grimmond S., Gustinch S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Saitana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,

Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukawa S.,  
RA Hara A., Hachizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sakai D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Kidney;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zesberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Saplenton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -|- FUNCTION: Cleaves peptide substrates on the N-terminus of arginine  
CC residues in dibasic pairs.  
CC -|- CATALYTIC ACTIVITY: Hydrolysis of polypeptides, preferably at  
CC Xaa-|-arg-Lys, and less commonly at Arg-|-Arg-Xaa, in which Xaa is  
CC not Arg or Lys.  
CC -|- COFACTOR: Zinc (Probable).  
CC -|- SIMILARITY: Belongs to peptidase family M16.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AK031548; BAC27445.1; -;  
CC EMBL; BC036128; AH36128.1; -;  
CC MGD; MGI:1201386; Nrnl.  
CC InterPro; IPR001431; Peptidase\_M16.  
CC InterPro; IPR007863; Peptidase\_M16\_C.  
CC Pfam; PF00675; Peptidase\_M16; 1.  
CC Pfam; PF05193; Peptidase\_M16\_C; 2.  
CC PROSITE; PS00143; INSULINASE; 1.  
CC Hydrolase; Metalloprotease; Zinc; Signal.  
CC SIGNAL 1 18  
CC POTENTIAL.  
CC CHAIN 19 1161 NARDILYSIN.  
CC DOMAIN 139 209 ASP/GLU-RICH (HIGHLY ACIDIC).  
CC DOMAIN 140 151 POLY-GLU.  
CC DOMAIN 152 165 POLY-ASP.  
CC DOMAIN 185 190 POLY-ASP.  
CC DOMAIN 193 196 POLY-ASP.  
CC METAL 244 244 ZINC (BY SIMILARITY).  
CC ACT\_SITE 247 247 BY SIMILARITY.  
CC METAL 248 248 ZINC (BY SIMILARITY).  
CC METAL 325 325 ZINC (BY SIMILARITY).  
CC SEQUENCE 1161 AA; 132890 MW; 21334221632A5122 CRC64;  
Query Match 3.7%; Score 111; DB 1; Length 1161;

Best Local Similarity 19.0%; Pred. No. 4.8;  
Matches 118; Conservative 64; Mismatches 190; Indels 250; Gaps 29;

QY 77 AFGKICTTSRLRLPGEVEVVGINCRWRLLCAEVECW--WCVHATHLHSGSLW 134  
Db 6 AVAACVTVGRKLRCEAGRELTALGRIBARG--LCEESSKFPFPLTWPGRNKAKSTCSCP 62  
QY 135 BILYQHSVRLKHKRRRRPRRPVGEN-----SDSSEEDHPAFCDVPVTO----- 177  
Db 63 DL-----QPNQDLSGRLARLGADESEEGSFSNVGDPILIKSPDPKQ 109  
QY 178 -----TG-----ASESDSGSGPSTRUSASG 198  
Db 110 YRYIKLQNGLOALLISLDSNVGKTGNATDEEBEEDDDDDDDDDDDDDSGAE 169  
QY 199 VQVVDNADSPGSDGSPSTRHSDSQPPADETTVHTDNDVDTLL-----D 247  
Db 170 IQDDDEGDFDEEEDDDDDHDDDD-----LENEELEELEERVEARKTTE 219  
QY 248 KESACALMYHVGQMDMLRAMDCEDLFDLLGI-----PE-----DVIATS 288  
Db 220 KQSAALCVGVG-----SFADPD--DLPLGLAHFLEHVMVMSGLKYPDENGDFAPLKK 269  
QY 289 QPGGD--TQASGVVTEGSAASAVGAGVEDVLAGAL-----EAQ 326  
Db 270 HGGSDNASTDCERTVFO-----FDVOKYFKKALDRWAOFHPLMIRDAIDREVE 320  
QY 327 NVAGEYVL-----EISDEVEDDAGLPPASRRRP--VVGEEFLWDDGPRRHERPTTRAI-RH 379  
Db 321 AVDSEYQIARSPDANRKEMLFG-----SLAPGHPMGKFFWNAETLKHEPKKNIDTH 374  
QY 380 KLASAYRVARPPVMTDLRGVEVFVFGPAMGLEVERK-----VFILCSQNPLA 430  
Db 375 ARLEEFNWR-----YSAHYMTLVVQSKETLDTLEKWTVEIFSQIP-- 415  
QY 431 DISHSLHSRKLRLVLPKPDNN--TGPGDVNLLAAVLR----- 468  
Db 416 -----NNG-----LPKNFSLHTDPTTAPNKLRYVVPTRKHALTITWALPQQQ 462  
QY 469 -----SPASGLVIVSLRGI--YKVLCKSTVLYHGNPNPKFGVIGLSRAVLDFV 519  
Db 463 HYRVKPLHYISLWVHGEGKGSILYLRKKWALFGNGE-----TGFEQNSYVSF 515  
QY 520 NVA-QYRIQGHKIKT--TVF 538  
Db 516 SISITLDEGEYEHFAHTVF 537

RESULT 9  
AEGP RAT STANDARD; PRT; 1216 AA.  
AC Q63191;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Apical endosomal glycoprotein precursor.  
OS Rattus norvegicus (Rat)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Intestinal epithelium;  
RX MEDLINE=95130529; PubMed=7829488;  
RA Speelman B.A., Allen K., Grounds T.L., Neutra M.R.,  
RA Kirchhausen T., Wilson J.M.;  
RT "Molecular characterization of an apical early endosomal glycoprotein  
RT from developing rat intestinal epithelial cells."  
RL J. Biol. Chem. 270:1583-1588(1995).  
CC -I- FUNCTION: Probably involved in the sorting and selective transport  
CC of receptors and ligands across polarized epithelia.  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
CC -I- TISSUE SPECIFICITY: Apical endosomal tubules of developing rat

CC intestinal epithelial cells.  
CC -I- SIMILARITY: Contains 3 LDL-receptor class A domains.  
CC -I- SIMILARITY: Contains 6 MAM domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
CC EMBL; L37380; AAA65200.1; -.  
CC PIR; A55620; A55620.  
CC InterPro; IPR002172; LDL receptor\_A.  
CC InterPro; IPR000998; MAM\_domain.  
CC Pfam; PF00057; ldl\_recept\_a; 1.  
CC Pfam; PF00629; MAM; 6.  
CC PRINTS; PR00261; LDLRECEPTOR.  
CC SMART; SM00137; MAM; 3.  
CC PROSITE; PS01209; LDLRA\_1; 2.  
CC PROSITE; PS00668; LDLRA\_2; 2.  
CC PROSITE; PS00740; MAM\_1; FALSE NEG.  
CC PROSITE; PS00660; MAM\_2; 6.  
CC Signal; Transmembrane; Glycoprotein; Repeat.  
CC SIGNAL 1 21  
CC FT CHAIN 22 1216 APICAL ENDOSONAL GLYCOPROTEIN.  
CC FT DOMAIN 22 1155 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSMEM 1156 1176 POTENTIAL.  
CC FT DOMAIN 1177 1216 CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 27 54 LDL-RECEPTOR CLASS A 1 (PARTIAL).  
CC FT DOMAIN 62 224 MAM 1.  
CC FT DOMAIN 229 269 LDL-RECEPTOR CLASS A 2.  
CC FT DOMAIN 268 427 MAM 2.  
CC FT DOMAIN 454 491 LDL-RECEPTOR CLASS A 3.  
CC FT DOMAIN 482 647 MAM 3.  
CC FT DOMAIN 654 813 MAM 4.  
CC FT DOMAIN 812 973 MAM 5.  
CC FT DOMAIN 972 1142 MAM 6.  
CC FT DISULFID 231 243 BY SIMILARITY.  
CC FT DISULFID 238 256 BY SIMILARITY.  
CC FT DISULFID 250 267 BY SIMILARITY.  
CC FT DISULFID 456 467 BY SIMILARITY.  
CC FT DISULFID 463 480 BY SIMILARITY.  
CC FT DISULFID 474 489 BY SIMILARITY.  
CC FT CARBOHYD 205 205 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 291 291 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 639 639 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC SQ SEQUENCE 1216 AA; 133776 MW; 64F3B28A7B61BA2E CRC64;  
Query Match 3.7%; Score 111; DB 1; Length 1216;  
Best Local Similarity 20.4%; Pred No. 5.2;  
Matches 99; Conservative 49; Mismatches 154; Indels 184; Gaps 25;

QY 30 RSTGFFLAARNVDVPLDSLOFFIDFKRECLSKGLHPRDLGSPITAFGKICTTSRLR 89  
Db 708 RQCGALLTRFPQVPVP-----KECLSFYH---LHGQFI---GTLCLAMRR-- 748  
QY 90 RLPGEVEVW---OGINCRRKLLCAEVEKWCWCHARTLHSGSSLWEII-----YQ 139  
Db 749 ---EGEEDTLWSRSGTHGNRWH-----QAWVTIH---HOLQSTKYQLFEGRLDGYH 796  
QY 140 HSVRLKHKRRPRFPVGENSDSSEEDHPAFCDVPVTOGTGAESDSDGEGPSTR-HSASG 198  
Db 797 GTMGDDNVAVFGPCWAAKRCFSFDS--CGF-----SPGDWGLWTRQNNASG 842  
QY 199 VQP-----VDDANADSPGS---GDGSP-----STRHSDSQPPADETTVHTDNDVDDL 243  
Db 843 LGPMGWIDHTTGTGAQGHYVMVVDTSPLLPGKHVASLTSEHPP----- 886











(Potential).  
-!- SIMILARITY: Contains 1 FtsK domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE014771; AN25210.1; -.  
HAWAP; MF\_01809; -; 1.  
InterPro; IPR003593; AAA ATPase.  
InterPro; IPR002543; FtsK SpoIIIE.  
InterPro; IPR008253; Marvel.  
Pfam; PF01580; FtsK SpoIIIE; 1.  
Pfam; PF01284; MARVEL; 1.  
SMART; SM00382; AAA; 1.  
PROSITE; PS00901; FTSK; 1.  
KW Chromosome partition; Cell division; ATP-binding; DNA-binding;  
Transmembrane; Complete proteome.  
FT TRANSMEM 73 92 POTENTIAL.  
FT TRANSMEM 102 124 POTENTIAL.  
FT TRANSMEM 136 158 POTENTIAL.  
FT TRANSMEM 168 190 POTENTIAL.  
FT TRANSMEM 192 214 POTENTIAL.  
FT DOMAIN 598 798 FTSK.  
FT NP BIND 615 622 ATP (POTENTIAL).  
SQ SEQUENCE 969 AA; 102650 MW; 50711A08044CB794 CRC64;  
  
Query Match 3.68; Score 107; DB 1; Length 969;  
Best Local Similarity 22.5%; Pred. No. 7.4;  
Matches 97; Conservative 42; Mismatches 142; Indels 150; Gaps 22;  
  
Qy 132 SLWEILYQHSVRLEKRRR-----PRPFV--GENSDSSSEDRP-----AFCDVP 174  
Db 205 SLLMITGHTYDLPEDARKAAKIQKPYPMQGETDGSASQFPNEVRVGDITLAD-- 262  
Qy 175 VTQGAESDSGDEGSPSTRHSAGVQPVDDANADSPGSGDEGPST-----RHSDSQPPP 228  
Db 263 ----GVPSHDGDDG-----SDNDQAGDARPSLFAFLGRKS----- 295  
Qy 229 ADETTVHTDNVEDTLTLDKESA-----CALMYHVGQEMDLRMAMCEDLFDLAGIPE 282  
Db 296 -----KTEDDKT-LDKYAADFPDRASQHGATATPTPV-----DPMTEIIG-AR 339  
Qy 283 DVATISQPG-----GDTASGVVTEGSIASAVGAGVED----- 316  
Db 340 TIASSSYDGRPHLSPPAPADADDGASRTRVITSGTVAMPGGGAVDPPWAPSAAQGT 399  
Qy 317 VYLAGALEAQNV-----GEYVLEISD-----EEVDDGAGLPASRRRVRVGE 359  
Db 400 VALAGAGAGVGAAGAGAAAAAATGAYAGADADGSGVGQGVNTGGQFNAT-----AGN 454  
Qy 360 FLWDGDPRRHERP-----TTRIRHRKLSAYYVARPPVMTIDRLGVEV-----FYFGREPA 411  
Db 455 DTDDANRPVQLPDLNLLTKGQPHAMETPANDVIRALTSTFEQNFNDKAVKVGFLGPSV 514  
Qy 412 MSLEVE-----RKVFILCSQNPDLADISHCSLHRSKGLRVLLPKPDNDNTG-----PG 458  
Db 515 TQTEVELGPGVKYEV-----TNLQRIAYAVASS--DVRILSPIPKGSAIGIEIPNEDR 567  
Qy 459 DVNLLAAVLRS 469  
Db 568 EIVHLGDVLR 578  
  
RESULT 14  
CLNN MOUSE  
ID CLNN MOUSE STANDARD; PRT: 1052 AA.  
AC Q8C5W0; Q91V71; Q91X77; Q91XT8; Q91XU9;  
DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Calmin.  
GN CLNN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI TaxID=10090;  
RN [1]\_TaxID=10090;  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4), SUBCELLULAR LOCATION,  
RP AND TISSUE SPECIFICITY.  
RC STRAIN=ICR; TISSUE=Brain, and Testis;  
RX MEDLINE=21280911; PubMed=11386753;  
RT Ishisaki Z., Takaiishi M., Furuta I., Huh N.-H.;  
RT "Calmin, a protein with calponin homology and transmembrane domains  
RT expressed in maturing spermatogenic cells";  
RL Genomics 74:172-179(2001).  
[2]  
RP SEQUENCE OF 98-1052 FROM N.A. (ISOFORM 1).  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=22354683; PubMed=12466851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru I., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Giasi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,  
RA Kongsaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Naglott D.R., Maltas L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Petrea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wushaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume K., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
[3]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=22557181; PubMed=12670712;  
RT Takaiishi M., Ishisaki Z., Yoshida T., Takata Y., Huh N.-H.;  
RT "Expression of calmin, a novel developmentally regulated brain  
RT protein with calponin-homology domains";  
RL Brain Res Mol Brain Res 112:146-152(2003).  
CC -!- SUBCELLULAR LOCATION: Type IV membrane protein (potential).  
CC Cytoplasmic. Isoforms 1 and isoform 4 show a reticular pattern in  
CC the cytoplasm.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=1; Synonym=Beta;  
CC IsoId=Q8C5W0-1; Sequence=Displayed;  
CC Name=2; Synonym=Delta;  
CC IsoId=Q8C5W0-2; Sequence=VSP\_007766, VSP\_007767;  
CC Note=Lacks the transmembrane domain;  
CC Name=3; Synonym=Gamma;  
CC IsoId=Q8C5W0-3; Sequence=VSP\_007768, VSP\_007769;  
CC Note=Lacks the transmembrane domain;  
CC Name=4; Synonym=Alpha;  
CC IsoId=Q8C5W0-4; Sequence=VSP\_007770;  
CC -!- TISSUE SPECIFICITY: Expressed in testis. Expressed during testis

maturation process and in maturing spermatids. In brain, it is expressed in neurons of the hippocampus, cerebral cortex, and thalamus, Purkinje cells, and also in the choroid plexus and ependymal cells. Expressed predominantly in dendrites and cell bodies of the neurons, but not in axons. The level of expression increases during the period of maturation of the mouse brain after birth.

-- SIMILARITY: Contains 1 actin-binding domain.  
-- SIMILARITY: Contains 2 calponin-homology (CH) domains.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR ENBL; AB047978; BAB59009.1; -
DR ENBL; AB059643; BAB59120.1; -
DR ENBL; AB059644; BAB59121.1; -
DR ENBL; AB059645; BAB59122.1; -
DR ENBL; AB059646; BAB59123.1; -
DR ENBL; AB059647; BAB59124.1; -
DR ENBL; AB059648; BAB59125.1; -
DR ENBL; AK077023; BAC36573.1; -
DR MG0; MG1;2136957; Clnm.
DR GO; GO:0005737; Cytoplasm; IDA.
DR InterPro; IPR001589; Actbind actnin.
DR InterPro; IPR001715; Calponin-like.
DR Pfam; PFC0307; CH; 2.
DR SMART; SM00033; CH; 2.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS50021; CH; 2.
DR Actin-binding; Repeat; Transmembrane; Alternative splicing.
KW ACTIN-BINDING; Repeat; Transmembrane; Alternative splicing.
FT TRANSMEM 1027 1047
  ANCHOR FOR TYPE IV MEMBRANE PROTEIN

```

FT				(POTENTIAL).	
FT	D0MAIN	1	288	ACTIN-BINDING.	
FT	D0MAIN	1	139	CH 1.	
FT	D0MAIN	187	288	CH 2.	
FT	VARSPLIC	922	927	NHSDS -> TVIPFL (in isoform 2).	
FT				/FtId=VSP 007766.	
FT	VARSPLIC	928	1052	Missing (in isoform 2).	
FT				/FtId=VSP 007767.	
FT	VARSPLIC	942	946	DHSY -> SFHLX (in isoform 3).	
FT				/FtId=VSP 007768.	
FT	VARSPLIC	947	1052	Missing (in isoform 3).	
FT				/FtId=VSP 007769.	
FT	VARSPLIC	966	996	Missing (in isoform 4).	
FT				/FtId=VSP 007770.	
FT	CONFLICT	170	170	S -> R (IN REF. 2).	
FT	CONFLICT	1012	1012		
FT	SEQUENCE	1052	AA;	117226 MW; A5AD7D3FF9A5EB6 CRC64;	
ST					

Query Match                 3.6%; Score 107; DB 1; Length 1052;  
 Best Local Similarity   20.8%; Pred. NO. 8.2;  
 Matches 105; Conservative 53; Mismatches 200; Indels 148; Gaps 23

Qy	20	DKVPLVKWLDRTGTFTLPAARNVDVPLDSLOFFIDFKRECLSKGLHPRDLLSPITAFG	79
Db	348	DKVFCVQDLSEPTG-FCFLDSAPSHKXSDSTEFVHEIIDVLOGSTKGTGSIAP-TPES	405
Qy	80	KICTT---SRRRLRIPGEEYEVVQINGCRRWLLCAEVKECMCVHARTHLHGSSL-VE	135
Db	406	SILSTFKDGRSNSLP-----VKK---TVHPADLHKDASCSD	441
Qy	136	ILYQHSVRLEKURRPRPRPFVGENSDSEEDHPAFCDVPVQTGAES-----EDSGD	189
Db	442	PFYSDFRPEGSPKATKL-----SKQDGHVSLAEVSKKKEQEARLVLEAASDKV	494
Qy	190	PSTRHSASGVQVDDANADSPGSGD-----EGPSTRHSDSOPP--PADETTVHTDVV	239

```

495 PE-----STV DGLDAVPDQAQPSODSSFCNGTGVESPSSQGEKGPFPSSPGDHTLL--ANST 548
240 EDDLTLDPKESACALMYHVGQEMDMLRAMCDEDLFDLLGIPE-----DVIATSQPG 291
549 ELKVQLLTVE-----PMDKEDYFECIPLKASKFNRDLVDPASTSQAP 590
292 GDTDASGVVTVTGSTAASAVGAVZEDVYLAGALTAQVAGEVYLE-----ISDEVDQ 343
591 GDPSHSEKTRGESEGSNHAEEKPKKSKSPRAETEAESRLEPKLPPPKDPQEQEQ 650
344 GAGLPPASRRPVRVGEFWDGPRHERPPTTRIRHKLRSAYYRVARPPVMITDLGVE 403
651 GHALPP-----ETPADKKPVYEKAKGKGRHHSBEEG-----683
404 VYVFGRPAMSLVERKVPILCSQN-----PLADISHSLC--HSRKGRLVLLPKPDNNNTG 456
684 -----EAGSGFSVACEEIPASPGSTSVSLETLRSHSEGLD-FKSPSPPLUSKIS 731
457 -----PGDVNL-----LAAVLSRFASG 473
732 VTPHDLFYVPHVPEVPLAAVLEAYAEAG 757

```

RESULT 15  
LDS DROME

ID	LDS DROME	STANDARD	PRT: 1061 AA.
LD	LD		
AC	AC	034739; G9VHY1;	
AC	AC	DT 1-FEB-1994	
DT	DT	10-OCT-2003 (Rel. 28, Created)	
DT	DT	10-OCT-2003 (Rel. 42, Last sequence update)	
DT	DT	15-MAR-2004 (Rel. 43, Last annotation update)	
DE	DE	DE Probable helicase lodestar.	

Q5 *Drosophila melanogaster* (Fruit fly)

OC Drosophila melanogaster. (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
PV [1]

RP SEQUENCE FROM N. A.

RX MEDLINE=92009170; PubMed=1916263;

RA Girdham C.G., Glover D.M.

RT "Chromosome tangling and breakage at anaphase result from mutations  
RT in *lodestar*, a *Drosophila* gene encoding a putative nucleoside  
RT triphosphate-binding protein.";

RL Genes Dev. 5:1786-1799(1991).

RN [2]

RP SEQUENCE FROM N.A.

```
RC S'TRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
```

RA Adams M.D.: Celniker S.E., Holt B.  
MEDHHS=20138008; PubMed=10731132;

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer R.A.

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.  
PA Abail T.E. 2000. A new species of the genus *Phrynosoma* from the  
PA Abail T.E. 2000. A new species of the genus *Phrynosoma* from the

RA ADRII J.F., AGDAYANI A., AH H.-J., ANDREWS-PIANKOCH C., BAIDWIN  
RA BALLEW P M BASU A BAYENDALE J BAYRAKTAROGU I BEAGLEY E

Barlow K.M., Basu A., Baxendale J., Bayraktarova E., Beasley D., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., RA

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chand

Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., RA

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Duvbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischman

RA Durbin K.J.; Evangelista C.C.; Ferraz C.; Ferreira S.; Fleischman  
RA Foster C.; Gabriellian A.E.; Garg N.S.; Gelbart W.M.; Glasser K.

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Liu  
Liu X., Mattei R., McIntosh T.C., McLeod M.P., McPherson D.

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

# THE UNIVERSITY OF CHICAGO

Query Match	3.6%;	Score 107;	DB 1;	Length 1061;
Best Local Similarity	22.7%;	Pred. No. 8.3;		
Matches	82;	Conservative 52;	Mismatches 149;	Indels 78; Gaps 17;
QY	158	ENSDSEED--HPAFCDVPVTQTG--AESDSGDE-----GPSTRHSAGSGVPV---D	203	
Db	82	QSEDESDSVRSANTKKKPGIGIPSDSEDEDELEQRLSFSSTRMSITGVAPQLSD	141	
QY	204	DANADSPGSGDEGFSRHSQSPPPADE-----TTVHTDNVEDDL-----TILDK	249	
Db	142	DSEIYSDEVQEGP-----TEAPTAAVVRVTTQFAGNIQNDLHSTIGAADESVLDD	195	
QY	250	SACALMHVQCEMDMLRAMCDEDLF---DLGIP-EDVIAISQGGDDTASGVTEGS	304	
Db	196	SGSDVILSNKPTPIELSTDDATNTKENMSGPPFERPSKLSRPSAGASVVKTSKN	255	
QY	305	IAASAVGAGVEDVYLAGALEAQNVAGEYLVLEIS-----DEEV---PDGAGLPASRRPV	356	
Db	256	LSQPTQAVLKQKTSPPAAPRSRIKESDQKVSVQVYVDEEMRLAEKRVQVSDAEKLF	315	
QY	357	VGSEFLWDDGPRRHERPTTR-----TERRKLSAYYRVARP-----PVMIT	397	
Db	316	VAHKLDPKGSQIKRIDITLRELAMDEQWTSALRVQOSNVPAVRVVKPTLNPRAPSIDT	375	
QY	398	---DRLG-----VEVFFYGFGRPAMSLVERVVFILCSQNPLADTSHSCLHSRKGRLVLPK	449	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 08:57:44 ; Search time 45 seconds  
(without alignments)  
3975.529 Million cell updates/sec

Title: US-10-656-499-2

Perfect score: 2993

Sequence: 1 MAGRLTWISFIVGALDSD.....QFDNVPLVIXLRSLRVTCTDD 567

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2993	100.0	567	12 Q99AM3	Q99am3 kaposi's sa
2	2977.5	99.5	566	12 Q9DUN1	Q9dun1 kaposi's sa
3	1620	54.1	313	12 Q40940	Q40940 kaposi's sa
4	828	27.7	152	12 Q9DUM9	Q9dum9 kaposi's sa
5	180.5	6.0	467	12 Q40941	Q40941 kaposi's sa
6	177.5	5.9	467	12 P88948	P88948 kaposi's sa
7	155.5	5.2	696	12 Q40939	Q40939 kaposi's sa
8	155.5	5.2	696	12 P88947	P88947 kaposi's sa
9	126.5	4.2	1373	10 Q9AX46	Q9ax46 oryza sativ
10	125.5	4.2	777	5 Q44122	Q44122 drosophila
11	125	4.2	777	5 Q9VL91	Q9vl91 drosophila
12	123	4.1	462	13 Q3DFH2	Q3dfh2 brachydanio
13	122	4.1	364	12 Q3WRN7	Q3wrn7 macaca mula
14	121.5	4.1	947	5 Q9BIA3	Q9bia3 caenorhabdi
15	121.5	4.1	1033	16 Q8D6Y3	Q8d6y3 vibrio vuln
16	121	4.0	361	12 Q9J2U5	Q9j2j5 macaca mula

17	120	4.0	253	12 Q9WRP1	Q9wrp1 macaca mula
18	120	4.0	998	12 Q9DWD5	Q9dwd5 rat cytoleg
19	119.5	4.0	1055	11 Q8CHG2	Q8chg2 mus musculu
20	119.5	4.0	1415	11 Q8OTFO	Q8otfo mus musculu
21	118	3.9	163	12 Q71294	Q71294 kaposi's sa
22	117.5	3.9	626	2 Q8KPN5	Q8kpn5 streptococc
23	117.5	3.9	1567	11 Q35243	Q35243 mus musculu
24	117	3.9	1316	4 Q75128	Q75128 homo sapien
25	116.5	3.9	487	13 Q8JFA9	Q8jfa9 oryzaia lat
26	116.5	3.9	506	11 Q8OV12	Q8ov12 mus musculu
27	116.5	3.9	565	13 Q8JIP0	Q8jip0 oryzaia lat
28	116	3.9	512	10 Q9LTH6	Q9lth6 arabidopsis
29	115.5	3.9	466	11 Q8C916	Q8c916 mus musculu
30	115.5	3.9	3436	12 Q66666	Q66666 equine herp
31	114.5	3.8	672	10 Q851R1	Q851r1 oryza sativ
32	114	3.8	691	4 Q9AQF7	Q9aqf7 homo sapien
33	112.5	3.8	748	10 Q8L734	Q8l734 arabidopsis
34	112.5	3.8	748	10 Q9FRT1	Q9ftr1 arabidopsis
35	111	3.7	1161	11 Q8R320	Q8r320 mus musculu
36	110.5	3.7	566	5 Q8T6T5	Q8t6t5 trypanosoma
37	110	3.7	974	5 Q9Y159	Q9y159 drosophila
38	110	3.7	975	5 Q8GB47	Q8gb47 drosophila
39	109.5	3.7	257	11 Q9CU64	Q9cu64 mus musculu
40	109.5	3.7	716	2 Q8RL15	Q8rl15 providencia
41	109.5	3.7	853	16 Q8ZUM5	Q8zum5 nitrosomona
42	109.5	3.7	1246	5 Q9NEE0	Q9nee0 leishmania
43	109	3.6	1026	5 Q9N6D9	Q9n6d9 leishmania
44	109	3.6	1664	10 Q9ZQK0	Q9zqk0 arabidopsis
45	109	3.6	4717	3 Q94248	Q94248 schizosacch

#### ALIGNMENTS

##### RESULT 1

Q99AM3 ID Q99AM3 PRELIMINARY; PRT; 567 AA.  
AC Q99AM3; Q99AM3 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-JUN-2001 (TREMREL. 24, Last annotation update)  
DE B-cell specific latent nuclear protein.  
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxID=37296;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20569372; PubMed=11119611;  
RA Rivas C., Thlick A.E., Parravicini C., Moore P.S., Chang Y.;  
RT 'Kaposi's sarcoma-associated herpesvirus LAN2 is a B-cell-specific  
latent viral protein that inhibits p53.';  
RL J. Virol. 75:429-438(2001).  
DR EMBL; AY008303; AAG50357.1; -;  
DR GO; GO:0005634; C:nucleus; IEA.  
KW Nuclear protein.  
SQ SEQUENCE 567 AA; 62663 MW; 7AF404225638D3DC CRC64;

Query Match 100.0%; Score 2993; DB 12; Length 567;  
Best Local Similarity 100.0%; Pred. No. 3.6e-228;  
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAGRLTWISFIVGALDSDKPYLVKWLDRSTGTFELAPARNNDVPLDSLOFFIDFKREC 60  
Db 1 MAGRLTWISFIVGALDSDKPYLVKWLDRSTGTFELAPARNNDVPLDSLOFFIDFKREC 60  
Qy 61 LSKGLPRDLGSPITAFGKICTTSRRRLRRLPGEEYEVVQGINCRWRLLCAEVKSCWC 120  
Db 61 LSKGLPRDLGSPITAFGKICTTSRRRLRRLPGEEYEVVQGINCRWRLLCAEVKSCWC 120  
Qy 121 VHARTLHSGSSWEILYQHSVLEKRRPRPRPFVGENSDSSEEDHPAFCDVPTQTGA 180  
Db 121 VHARTLHSGSSWEILYQHSVLEKRRPRPRPFVGENSDSSEEDHPAFCDVPTQTGA 180

QY 181 ESDSGDEGSTRHSGVQPVDDANADSPGSGDEGSTRHSDSQPPADTTVHTDNVE 240  
 DB 181 ESDSGDEGSTRHSGVQPVDDANADSPGSGDEGSTRHSDSQPPADTTVHTDNVE 240  
 QY 241 DLTLLDKESACALMVHVGQEMDLVRAMCDEDLFLLGIPEDVIATSPGSGDTHASGV 300  
 DB 241 DLTLLDKESACALMVHVGQEMDLVRAMCDEDLFLLGIPEDVIATSPGSGDTHASGV 300  
 QY 301 TEGSIAASAVGAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPPASRRRPVVGGEF 360  
 DB 301 TEGSIAASAVGAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPPASRRRPVVGGEF 360  
 QY 361 LWDGPRRHERPTTRIRHRKLSAYYRVARPPVMTDRLGVEVYFGRPAMSLVERKV 420  
 DB 361 LWDGPRRHERPTTRIRHRKLSAYYRVARPPVMTDRLGVEVYFGRPAMSLVERKV 420  
 QY 421 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 480  
 DB 421 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 480  
 QY 481 SGIYVKNLCKSVLYHGNNPPKFGVICGLSSRAVLDVFNVAQYRIQGHIEHKKTTVFIG 540  
 DB 481 SGIYVKNLCKSVLYHGNNPPKFGVICGLSSRAVLDVFNVAQYRIQGHIEHKKTTVFIG 540  
 QY 541 GDPTSABQFDMVPLVILKRLRSVTCDD 567  
 DB 541 GDPTSABQFDMVPLVILKRLRSVTCDD 567  
 RESULT 2  
 Q9DUN1 PRELIMINARY; PRT; 566 AA.  
 ID Q9DUN1  
 AC Q9DUN1  
 DT 01-JAN-2001 (TrEMBLrel. 16, Created)  
 DT 01-VAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DE Similarity to 16.  
 DE Interferon regulatory factor 3.  
 GN VIRF-3.  
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Rhadinovirus.  
 OX NCBI\_TaxID=37296;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20392178; PubMed=10933732;  
 RA Lubyova B., Pitha P.M.;  
 RT "Characterization of a novel human herpesvirus 8-encoded protein,  
 RT VIRP-3, that shows homology to viral and cellular interferon  
 RT regulatory factors." (2000).  
 RL J. Virol. 74:8194-8201 (2000).  
 DR EMBL; AF157602; AAF23608.2; -.  
 SQ SEQUENCE 566 AA; 62507 MW; 822C557407A8C84B CRC64;  
 Query Match 99.5%; Score 2977.5; DB 12; Length 566;  
 Best Local Similarity 99.8%; Pred. No. 6.1e-227;  
 Matches 566; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 1 MAGRLTWISFIVGALSDSKYPLVKWLDRTSGTFLAPARNNDVPLDSLOFFIDFKREC 60  
 DB 1 MAGRLTWISFIVGALSDSKYPLVKWLDRTSGTFLAPARNNDVPLDSLOFFIDFKREC 60  
 QY 61 LSKGLHPRDLGSPITAFGKICTTERRLRLLPGEEYVYVQGINCRWELLCAEVKECWC 120  
 DB 61 LSKGLHPRDLGSPITAFGKICTTERRLRLLPGEEYVYVQGINCRWELLCAEVKECWC 120  
 QY 121 VHARTHLHSGSISWELLYQHSVRLKRRRRPRPFGVNSDSSSEEDHPAFCDVPVTQGA 180  
 DB 121 VHARTHLHSGSISWELLYQHSVRLKRRRRPRPFGVNSDSSSEEDHPAFCDVPVTQGA 180  
 QY 181 ESDSGDEGSTRHSGVQPVDDANADSPGSGDEGSTRHSDSQPPADTTVHTDNVE 240  
 DB 180 ESDSGDEGSTRHSGVQPVDDANADSPGSGDEGSTRHSDSQPPADTTVHTDNVE 239

QY 241 DLTLLDKESACALMVHVGQEMDLVRAMCDEDLFLLGIPEDVIATSPGSGDTHASGV 300  
 DB 240 DLTLLDKESACALMVHVGQEMDLVRAMCDEDLFLLGIPEDVIATSPGSGDTHASGV 299  
 QY 301 TEGSIAASAVGAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPPASRRRPVVGGEF 360  
 DB 300 TEGSIAASAVGAGVEDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPPASRRRPVVGGEF 359  
 QY 361 LWDGPRRHERPTTRIRHRKLSAYYRVARPPVMTDRLGVEVYFGRPAMSLVERKV 420  
 DB 360 LWDGPRRHERPTTRIRHRKLSAYYRVARPPVMTDRLGVEVYFGRPAMSLVERKV 419  
 QY 421 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 480  
 DB 420 FILCSQNPDLADISHSCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLR 479  
 QY 481 SGIYVKNLCKSVLYHGNNPPKFGVICGLSSRAVLDVFNVAQYRIQGHIEHKKTTVFIG 540  
 DB 480 SGIYVKNLCKSVLYHGNNPPKFGVICGLSSRAVLDVFNVAQYRIQGHIEHKKTTVFIG 539  
 QY 541 GDPTSABQFDMVPLVILKRLRSVTCDD 567  
 DB 540 GDPTSABQFDMVPLVILKRLRSVTCDD 566  
 RESULT 3  
 O40940 PRELIMINARY; PRT; 313 AA.  
 ID O40940  
 AC O40940  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Similarity to K10.  
 DE Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Rhadinovirus.  
 OX NCBI\_TaxID=37296;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97296220; PubMed=9151804;  
 RA Neipel F., Albrecht J.C., Fleckenstein B.;  
 RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus  
 RT human herpesvirus 8: determinants of its pathogenicity";  
 RL J. Virol. 71:4187-4192 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Neipel F., Albrecht J.-C., Ensaeer A., Huang Y.-Q., Li J.J.,  
 RA Friedman-Kien A.E., Fleckenstein B.;  
 RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";  
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EXBL; U93872; AAB62648.1; -.  
 SQ SEQUENCE 313 AA; 34319 MW; 0069EA68F69500BE CRC64;  
 Query Match 54.1%; Score 1620; DB 12; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 7e-120;  
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 255 MYHVGQEMDLVRAMCDEDLFLLGIPEDVIATSPGSGDTHASGVVTEGSIASAVGAGV 314  
 DB 1 MYHVGQEMDLVRAMCDEDLFLLGIPEDVIATSPGSGDTHASGVVTEGSIASAVGAGV 60  
 QY 315 EDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPPASRRRPVVGGEVWDGPRRHERPTT 374  
 DB 61 EDVYLAGALEAQNVAGEVYVLEISDEEVDGAGLPPASRRRPVVGGEVWDGPRRHERPTT 120  
 QY 375 RRIRHRKLSAYYRVARPPVMTDRLGVEVYFGRPAMSLVERKVFIICSNPLADISH 434  
 DB 121 RRIRHRKLSAYYRVARPPVMTDRLGVEVYFGRPAMSLVERKVFIICSNPLADISH 180  
 QY 435 SCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLRSGIYVKNLCKSV 494  
 DB 181 SCLHSRKGRLVLLPKPDNNNTGPGDNNLLAAVLSRSPASGLVIVSLRSGIYVKNLCKSV 240

QY 495 YHGNPNPKFGVTCGLSSRAVLDFVNAQYRIQGHHEIKTTVFVIGDPTSAEQDMVPL 554  
DB 241 YHGNPNPKFGVTCGLSSRAVLDFVNAQYRIQGHHEIKTTVFVIGDPTSAEQDMVPL 300

QY 555 VIKLRLSVTCDD 567  
DB 301 VIKLRLSVTCDD 313

RESULT 4  
Q9DUM9 PRELIMINARY; PRT; 152 AA.  
AC Q9DUM9;  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
DE ORF K10.6.  
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
NCBI\_TaxID=37296;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=20392178; PubMed=10933732;  
RA Lubyova B., Pitha P.M.;  
RT "Characterization of a novel human herpesvirus 8-encoded protein,  
RT vIRF-3, that shows homology to viral and cellular interferon  
RT regulatory factors.";  
RL J. Virol. 74:8194-8201(2000).  
DR EMBL; AF254765; AAF98389.1; --  
SQ SEQUENCE 152 AA; 17801 MW; 9492017B2BCA7076 CRC64;

Query Match 27.7%; Score 828; DB 12; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.2e-57;  
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGRRITWISERTVGLSDSKYPLKWLDRSTGTFILAPARNDVPLDSLOPFIDFKREC 60  
DB 1 MAGRRITWISERTVGLSDSKYPLKWLDRSTGTFILAPARNDVPLDSLOPFIDFKREC 60

QY 61 LSKGLHPRDLGSPITAFGKICTTSRRRLRPGEEVYVQGCINCRWRLLCAEVKECWWC 120  
DB 61 LSKGLHPRDLGSPITAFGKICTTSRRRLRPGEEVYVQGCINCRWRLLCAEVKECWWC 120

QY 121 VHARTHLHGSSLWEILYOHVSRLEKRRRPR 152  
DB 121 VHARTHLHGSSLWEILYOHVSRLEKRRRPR 152

RESULT 5  
O40941 PRELIMINARY; PRT; 467 AA.  
AC O40941;  
DT 01-JAN-1998 (T-EMBLrel. 05, Created)  
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE ORF K11.  
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
NCBI\_TaxID=37296;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97296220; PubMed=9151804;  
RA Neipel F., Albrecht J.C., Fleckenstein B.;  
RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus  
RT human herpesvirus 8: determinants of its pathogenicity?";  
RL J. Virol. 71:4187-4192(1997).  
[2]  
RN SEQUENCE FROM N.A.  
RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,  
RA Friedman-Kien A.E., Fleckenstein B.;

"The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U93872; AB62647.1; --  
SQ SEQUENCE 467 AA; 50467 MW; 9E72A0155F325SAC CRC64;

Query Match 6.0%; Score 180.5; DB 12; Length 467;  
Best Local Similarity 21.7%; Pred. No. 9.9e-06;  
Matches 106; Conservative 72; Mismatches 186; Indels 125; Gaps 22;

QY 167 HPAFCDVP-VTQTGAESDSDGEGP-STRSASGVQ-----PVDDANADSPGSGDEGP 217  
DB 2 HSLFFEPFSGFGSSQSSLLAPDSFRSTSQVQGPLVHTPTD---LCLPFGGLPSP 58

QY 218 STRHSDSQ---PPPADETTV-----HTDNVEDDLTLDKESACA-LMYHVGQEM 262  
DB 59 VIFPHETQGLLAPAGSQSQTPFSPGEPVPSHVSGLDCLPMDHIEGCLLDLSDVQCEL 118

QY 263 -DM--LNRAMCD-----EDLFD 276

DB 119 PDLDGLGELLCETASPOGPMQSEGEGESTESVSLPATHPLESSAPGASVMGSGQLPDP 178

QY 277 LLGIPEDVIATSPGGTDASGVVVTGSS-----IAASAVAGVEDVYLAGALE 324

DB 179 LGDLSELLCETASPOGPMQSEG-GEESTESVSLPATHPLESSAPGASV---MGSSFQ 233

QY 325 AQNVAGYV-----LEISDEVDGAGLPPASRRRPVVGFLWDDGRRRHERPTRR--I 377

DB 234 ASDNVDFIDICPLCRDRED-----QEKADQTFYWGSDMRPKVLTATOSVA 284

QY 378 RHKRLRSAYYRVARPPVMTDRLGVEVFYFGRPAMSLVERKVFILCSQNLADISHSCL 437

DB 285 AYLKSKQAIYKVGKLVPLV-----VEVYFGEKVKTHFDLTGGIVICSOVPEASPEHC- 339

QY 438 HSRKGLRVLLPKP-----DNNTGPGVNNLLAAVLRFSAGSLVIVSRGIVYKVLCKST 492

DB 340 QTVPYKCLLPRTAHCSVDANRT-----LEQTLDRFSMGVVAIGTNMGIKGLLEY 392

QY 493 VLHGNPNPKFGVTCGLSSR-AVLDFVNAQYRIQGHHEIKTTVFVIGDPTSAEQDM 551

DB 393 AYFGNARSRRIGKCRPLSHRHEIQQAFDVEREN-REPEGSRYASLFLGRRSPEDYDNDH 451

QY 552 VPLVIKRL 560

DB 452 YPVLHIYL 460

RESULT 6  
P88948 PRELIMINARY; PRT; 467 AA.  
AC P88948;  
DT 01-MAY-1997 (T-EMBLrel. 03, Created)  
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE ORF K11.  
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
NCBI\_TaxID=37296;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97094384; PubMed=8939871;  
RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;  
RT "Molecular mimicry of human cytokine and cytokine response pathway  
RT genes by KSHV.";  
RL Science 274:1739-1744(1996).  
[2]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=97121480; PubMed=8962146;  
RA Russo J.J., Bochenky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,  
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;  
RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus  
RT (HHV8).";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).



[3]  
SEQUENCE FROM N.A.  
RU Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,  
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U75698; AAC57142.1; -  
SQ SEQUENCE 467 AA; 5036 MW; 9E72A7789F5325AC CRC64;  
Query Match 5.9%; Score 177.5; DB 12; Length 467;  
Best Local Similarity 21.7%; Pred. No. 1.7e-05;  
Matches 106; Conservative 71; Mismatches 187; Indels 125; Gaps 22;  
QY 167 HPACDVP-VTQTAESEDSGDEGP-STRHSAAGVQ-----PVDANADSPGSGDEGP 217  
Db HSLFEEPEEPGFGSSGSSILLAPDSRPSTQVQGLVHVTDT---LCLPTGGLPSP 58  
QY 218 STRSDSQ---PPRADETIV-----HTDNVEDDITLLDKESACA--LMTYHVGQEM 262  
Db VIFPHETQGLLAPAGQSQTFPSPEGPVPSHVGDDCLPMDVHIEGCLLDLLSDVQQL 118  
QY 263 -DM--LMRAMCD-----  
Db 119 PDLGDLGELLCTASPOGPMQSEGEGSTESVSLPATHPLESSAPCASVMGSGQLPD 178  
QY 277 LLGIPEDVIATSQPGDITDASGVTEGS-----IAASAVGAGVEDVYLAGALE 324  
Db 179 LGDLSSELLCETASQGPQMSEG-GEESTESVSLPATHPLESSAPCASV-----MGSSFFQ 233  
QY 325 AQNVAGEYV-----LEISDEVDGAGLPPASRRRPVVGDFLWDGPRRHERPTTR--I 377  
Db 234 ASDNVDDFDICPLCLDRDRVED-----QEKADQTFYWGSDMRPKVLTATQSA 284  
QY 378 RHRKLRSAYYVARPPVPMITDLGVEYFGRPAWSLEVERKVFILCSQNPDLADISHSL 437  
Db 285 AYLKSKQAIYKGVKLVV-----VEVYFGEKVKTHFDLTGGIVICSQVPEASPEHIC- 339  
QY 438 HSRKGLRVLLPKP-----DDNNTGPGDNNLAAVLRSPFASGLVIVSLRSYVKNLCKST 492  
Db 340 QTVPPYKCLLPTAHCSVDANRT-----LEQTLDRFSMGVVAIGTNGIFLKGLEY 392  
QY 493 VLYHGNNPKFEGVICGLSSR-AVLVDVNVAYRIQGHIEKTKTTFVIGDPTSAEQFDM 551  
Db 393 AIFVGNASRRIGKCRPLSHRHEIQAPDVERHN-REPEGSRYASFLGRRPPEYDSH 451  
QY 552 VPLVILKRL 560  
Db 452 YPVLHIVL 460  
PRELIMINARY; PRT; 696 AA.  
O40939  
AC O40939;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE ORF K10.  
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxID=37296;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97296220; PubMed=9151804;  
RA Neipel F., Albrecht J.C., Fleckenstein B.;  
RT "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus  
human herpesvirus 8: determinants of its pathogenicity?";  
RL J. Virol. 71:4187-4192(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,  
RA Friedman-Kien A.B., Fleckenstein B.;  
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U93872; AAB62660.1; -  
SQ SEQUENCE 696 AA; 74036 MW; C0B46CCD3934F4A7 CRC64;  
Query Match 5.2%; Score 155.5; DB 12; Length 696;  
Best Local Similarity 22.9%; Pred. No. 0.0016;  
Matches 103; Conservative 58; Mismatches 192; Indels 97; Gaps 22;  
QY 151 PRPFVGENSDSSESDHFAFCDPVPTQTGAESSEDSGDEGPSTRHSAAGVQVDDANADSP 210  
Db 302 PPTPEVAERQEPSSSGIPVVCQ-----GDNMATGYRRVTTSAGALEVE-IIDLTGDS- 352  
QY 211 GSGDEGPSTRHSDSQPPPADETTVTDNVEDDITLLDKESACALMYHVGQEMDMLRMC 270  
Db 353 ----DTPSTVA-STFLPVSGPRVFPQTV---LYSAPFPAVNPVSHLPTLELR-RECV 403  
QY 271 DEDLFDLIGIPEDVIATSQPGDITDASGVTEGSIAASAVGAGVEDVYLAGALEAQNVA 329  
Db 404 PGS-----GERPRVPLVSTYAGDRYAVGVGPEQSLVPPPLGL-----PLTSLNLQ 449  
QY 330 GEYVL-----EISDEVDGAGLPPASRRRPVVGDFLWDGPRRHERPTTRIR 378  
Db 450 GEDICTWEELGNILSELQEE-----PSSTRQAT-----DR--RPRSRSPH 490  
QY 379 HRLKRSAYYVARPPVPMI-----TDLRG--VEVYFVG--REPMSLEVERKVFILCSQNP 428  
Db 491 GRTPVSHSGPEKPSKMFDPDPSQSRVSFVVEIFVYGNLRTLRREGDAGEAMLCSW-P 549  
QY 429 LAD-ISHCLSHSRKGL-----RVLLPKPDNNTGPGDNNLAAVLRSPFASGLVIVSLRS 483  
Db 550 VGDTLGHLQCQSVPELLRPLRTVPSPE-----QMEILNRFEGLGHPFIFCSMSGI 602  
QY 484 YVKNLCKSTVLYHGNNPKFEGVICGLSSRAVLVDVNVAYRIQGHIEKTKT----- 535  
Db 603 YSRNATQVEGMVFNPNRKYERILRSFSPVQQLFNTRY-LATTAIPOTPLSVNVPVT 661  
QY 536 --TVFIGDPTSAEQFDMVPLVILKRLRSV 563  
Db 662 CGTVFEGASPTENFQNVPLVILKIFIGSI 691  
RESULT 8  
P88947  
ID P88947  
AC P88947; PRELIMINARY; PRT; 696 AA.  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE ORF K10.  
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxID=37296;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97094384; PubMed=8939871;  
RA Moore P.S., Boshoff C., Weiss R.A., Chang Y.;  
RT "Molecular mimicry of human cytokine and cytokine response pathway  
genes by KSHV";  
RL Science 274:1739-1744(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97121480; PubMed=8962146;  
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,  
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;  
RT "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus  
(HHV8)";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,  
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;  
RT "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";

SQ SEQUENCE 1373 AA; 150143 MW; 7C1A301D051A9B31 CRC64;  
 Query Match 4.2%; Score 126.5; DB 10; Length 1373;  
 Best Local Similarity 23.1%; Pred. No. 0.82;  
 Matches 99; Conservative 57; Mismatches 148; Indels .125; Gaps 23;  
  
 QY 144 LEKHRRPRRPVGV-----ENSQSS--EEDHP-----AFCDVPVTQT 178  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 67 ISEKRNKRKYFFNGIEYDNSSFRNSESSTPRESRPRRLVALPLSLIFAAATDDD 126  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 179 GAESDSGDEG--PSTR-----HSASG--VQPVDANADSPSGSDGPGSTR----- 220  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 127 AVDADDADAGHWPERRRPOHTADAGSHRTDDAADADDAG-HWPHRCERRLALQA 185  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 221 HSDSQPPPADE---TTVHTNDVEDDLTLDKSACALMTHVGQMDMLVRAMCEDELFDL 277  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 186 AASPVAADKYSTSLATEDITEWMTEDDRSE---RYRDHQKGDTLMAAN-----NL 236  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 278 L-----GIPEDVIATSPGGDTDASGVVTEGSIAA-----SAVGAGVEDVYL 319  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 237 LVGAAPREAREAGDAASKLAAL--GTGLAAVSTATALAAPEAVTGGFAANTYY 294  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 320 AGALEAQNVAGEVLE----ISDEEVDDGAGLPAPSRPRPVGEFJWDGP-----FRHER 371  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 295 NLALAVTFLAGVTLCALVWSDK-----PAARRRAAGKKLLYAAIPLIVATRHR 345  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 372 PT-----TRIRHR-----KLRSNYRVAPPVMITDRIGVEVFYFGR 409  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 346 RSTGGGAALDLERNVIARCNSGDIYPYPATSTHALAAPTSLWHRRLG-----HLGR 401  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 410 PAMSLEVERKVILCSQNPLADISHSLRSKRLRVLLPKPDNNNTGPDVNLLAAVLRS 469  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 402 EALS-KLRSSVISCTKDPLLCHAC---QLGHHTRLPFSSSSSRASNFDLIHCDL-- 455  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 QY 470 PASGLVIVS 478  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
 Db 456 WTSPIVSVS 464  
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  
  
 RESULT 10  
 ID Q44122 PRELIMINARY; PRT; 777 AA.  
 AC Q44122;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 Dt Zinc finger 30C  
 GN ZF30C OR CG3998.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 RN NCBI\_TaxID=7227;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EYE/ANTENNAL DISC;  
 RA Cultured T., Gaul U.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF032275; AAB87630.1; -  
 DR FlyBase; FBgn0022720; zf30C.  
 DR InterPro; IPR007087; znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 13.  
 DR SMART; SM00355; znf\_C2H2; 12.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 10.  
 DR PROSITE; PS01517; ZINC\_FINGER\_C2H2\_2; 11.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 777 AA; 88280 MW; D1C69524E9E73CA4 CRC64;  
  
 Query Match 4.2%; Score 125.5; DB 5; Length 777;  
 Best Local Similarity 22.6%; Pred. No. 0.45;  
 Matches 88; Conservative 39; Mismatches 108; Indels 155; Gaps 23;

```
DB 62 RHCDARFCHBELTOHAKDEHGVTVGAVAGQERNAF--VCEKC-----GAEKYQAYR 112
QY 104 CRRWELLCAEVK-----ECWVCVHARHLHSGSLWELLYOHSVRLEKHRR----- 149
DB 113 -RHCTKCGEKLPRSESRPMECKCY---TRFSSAN-----LSKHRSRPT 157
QY 150 --RPRRPVVG-----ENSDSEEDHPAPCDVPVVTQTGAESDSGDEGPST 192
DB 158 CQPEYDSFGSDGKMKHKAFAKDKRNRSDDED-----TTSEESDSDDIPLA 207
QY 193 RHSASGVQPDADANADSPGSDGEGSTRHSQ-----PPADETIVHTDNVEDD 242
DB 208 SRLTKLKQ--ESQNSD---SGDECPDEPNNSDDADAGFQLPFA---MVKVEAFDEE 260
QY 243 LTLDKESACALMYHVGQEMDLNRMAMCDEDLFLL-----GI-PEDVIATSGGGTDA 296
DB 261 ----DFEYQDASMYVKTSTDFISNE--KOKLLDVLNNEGDLKPFPSLSLVEQ----- 307
QY 297 SGVVTGSIASAVGAGVEDVYLAGALEAQNVAGEYVLEISDEVEDDAGAGLPASRRPV 356
DB 308 -----GAGILDE--IAAVPLVEAEDVLELRGHQMEK-----PP----- 340
QY 357 VGEFLWDDGPRRHRPTTRRI-----RHRK 381
DB 341 -----GPRKGRGPRPKEKIPVVKRYK 362

RESULT 11
Q9VL91 ID PRELIMINARY; PRT; 777 AA.
AC Q9VL91
CT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Zf30C protein (Lb23102p).
GN ZF30C OR C63998.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RX MEDLINE=2019006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Rogers Y.-H.C., Blazej R.G., Champagne B.D.,
RA Abrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy L., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

```
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spue E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassaman D.A., Weinstock G.M., Weissbach J., Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Rao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2000).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Murgall C.J.,
RA Nunoo J., Pacle B.J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF503626; AAF52805.1; -
DR EMBL; AY051681; AAK93105.1; -
DR FlyBase; FBgn002720; zf30C.
DR InterPro; IPR007087; Znf C2H2.
DR Pfam; PF00096; zf-C2H2; 13.
DR SMART; SM00355; Znf C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 10.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 777 AA; 88347 MW; ED12A39DA882B994 CRC64;

Query Match 4.2%; Score 125; DB 5; Length 777;
Best Local Similarity 23.5%; Pred. NO. 0.49; Gaps 20;
Matches 80; Conservative 35; Mismatches 94; Indels 132;

QY 93 GBEYEVVQGINCRRLWLLCAEVK-----ECWVCVHARHLHSGSLWELLYOHSVR 143
DB 102 GAEYKQAYR-RHCRTKCGEKLPRSESRPMECKCY---TRFSSAN----- 146
QY 144 LSKHRR-----RPRRPVVG-----ENSDSEEDHPAPCDVPVVTQTGA 181
DB 147 LSKHRSRPTCGQPEYDSFGSDGKMKHKAFAKDKRNRSDDED-----TTSEE 196
QY 182 SEDSGEGSTRHSASGVQPDADANADSPGSDGEGSTRHSQ-----PPADE 231
DB 197 SEDSDDDIPLASRLTKLKQ--ESQNSD---SGDECPDEPNNSDDADAGFQLPFA-- 250
QY 232 TVVHTDNVEDDLTLDKESACALMYHVGQEMDLNRMAMCDEDLFLL-----GI-PEDVI 285
DB 251 -MVKVEAFDEE---DFEYQDASMYVKTSTDFISNE--KOKLLDVLNNEGDLKPFPSL 303
QY 286 ATSQPGGDTASGVVTGSIASAVGAGVEDVYLAGALEAQNVAGEYVLEISDEVEDDGA 345
DB 304 KVEQ-----GAGILDE--IAAVPLVEAEDVLELRGHQMEK-- 338
QY 346 GLPPASRRRPVVGELWDDGPRRHRPTTRRI-----RHRK 381
DB 339 --PP-----GPRKGRGPRPKEKIPVVKRYK 362

RESULT 12
Q9DFH2 ID PRELIMINARY; PRT; 462 AA.
AC Q9DFH2;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE HMG box transcription factor Sox9a.
GN SOX9A.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
```

OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chiang E.F., Pai C.-I., Van Y.-L., Postlethwait J., Chung B.-C.;  
 RT "Two Sox9 Genes on Duplicated Zebrafish Chromosomes: Overlapping  
 RT Expression Sites in Embryos but Distinct Expression in Adults";  
 RL Submitted (Jun-2000) to the EMBL/GenBank/DBJ Databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zeng L., Miller C.T., Nissen R., Singer A., Liu D., Kirn A.,  
 RA Draper B., Willoughby J., Morcos P.A., Amsterdam A., Chung B.C.,  
 RA Westerfield M., Hafler P., Hopkins N., Kimmel C., Postlethwait J.H.;  
 RT "A zebrafish sox9 gene required for cartilage morphogenesis";  
 RL Development 129:5065-5079(2002).  
 DR EMBL; AF277096; AAC09814.1; -  
 DR EMBL; AY090034; AM13695.1; -  
 DR EMBL; AY090035; AM13696.1; -  
 DR HSP; Q05066; IHRX;  
 DR ZFIN; ZDB-GENE-001103-1; sox9a.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000910; HM12\_box.  
 DR Pfam; PF00505; HM12\_box; 1.  
 DR SMART; SM00398; HM12; 1.  
 DR PROSITE; PS01118; HM12\_BOX\_2; 1.  
 SQ SEQUENCE 462 AA; 51333 MW; 78D10994FB67FAFC CRC64;  
 Query Match 4.1%; Score 123; DB 13; Length 462;  
 Best Local Similarity 21.9%; Pred. No. 0.34;  
 Matches 93; Conservative 40; Mismatches 134; Indels 158; Gaps 23;  
 QY 47 LSLQFPIPKRECLSKGLHP-----ROLLSPITAFKICTTTRRLRLPGBE----- 95  
 DB 4 LDPYKMTQEQKCLSDAFSPMSSEDSAGSPSPASGSDTENTR-----PAENSLLAADGT 59  
 QY 96 -----YEVVQGINCRWELLCAYKE-----CW 118  
 DB 60 LGDFKDEEDKFPVCIREAVSQVLEKYD---WTLVPMFVRVNGSSKNKPHVPRMFAFW 116  
 QY 119 WCYHAR-----THLHSG-----SSLWEILYQHSVRLEKRRPRPFVGENSD---S 162  
 DB 117 WAQARARLADQYPLHNAELSKTLGLWRLNE-----VEK-----RPFVEAERLRVQ 166  
 QY 163 SEEDHAPACDVP-----VTCGAESEDSGDE---GPSTRHSAGVOPVDANADSPGS--- 212  
 DB 167 HKXDPDYKYQPRRRKSVKNGSESEDSGSETHISPNALFKAL-----QOARDSPASMN 219  
 QY 213 GD-EGPSTRHSQPPPADETTVHTDNVEDDLTLDKESACALMYHVGQEMDMLRAMCD 271  
 DB 220 GEVHSFSEHSQSGQGPPTPTTKTDTQPKADL--KREARPLQENTGRPLSINFQ---D 274  
 QY 272 EDLFDLIGIPEDVIATISQ-----PGDPTDASGVVTEGSTAASAVGAGVEDVYLAGA 322  
 DB 275 VDIGEL---SSDVETFDVNEFDQYLPFNHQNQ-----PYAGG---YAAWM 315  
 QY 323 LEQNVAGEVLEISDEVDGAGLPPASRRRPPVGEFLWDDGPRRHERPTTTRIRHRKL 382  
 DB 316 TKFQN-----GSPQSSQLTEJ-----NPAPDQPRTHIKTEQL 349  
 QY 383 RSAYY 387  
 DB 350 SPSHY 354  
 RESULT 13  
 Q9WRN7 PRELIMINARY; PRT; 364 AA.  
 ID Q9WRN7  
 AC Q9WRN7  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE R13 (VIRP).  
 OS Macaca mulatta rhadinovirus 17577, and  
 OS Macaca mulatta rhadinovirus 26-95.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Rhadinovirus.  
 OX NCBI\_TaxID=83534, 119193;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Macaca mulatta rhadinovirus 17577;  
 RX MEDLINE=99174001; PubMed=10074154;  
 RA Seales R.P., Berguam E.P., Axthelm M.K., Wong S.W.;  
 RT "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with  
 RT similarity to Kaposi's sarcoma-associated herpesvirus/human  
 RT herpesvirus 8";  
 RL J. Virol. 73:3040-3053(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Macaca mulatta rhadinovirus 26-95;  
 RX MEDLINE=20173730; PubMed=10708456;  
 RA Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,  
 RA Desrosiers R.C.;  
 RT "The primary sequence of rhesus monkey rhadinovirus isolate 26-95;  
 RT sequence similarities to Kaposi's sarcoma-associated herpesvirus and  
 RT rhesus monkey rhadinovirus isolate 17577";  
 RL J. Virol. 74:3388-3398(2000).  
 DR EMBL; AF083501; AAF21391.1; -  
 DR EMBL; AF210726; AAF60043.1; -  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001346; IRF.  
 DR Pfam; PF00605; IRF; 1.  
 DR SMART; SM00348; IRF; 1.  
 SQ SEQUENCE 364 AA; 41189 MW; DBA48195C39DD018 CRC64;  
 Query Match 4.1%; Score 122; DB 12; Length 364;  
 Best Local Similarity 22.5%; Pred. No. 0.29;  
 Matches 82; Conservative 43; Mismatches 127; Indels 112; Gaps 19;  
 QY 246 LDKESACALMYHVGQEMDMLRAMCDDELDLIGIPEDVIATISQPGDPTDASGVVTEGSI 305  
 DB 57 LDRLECGRQHVSECRKRLRLVLRNAGFE---QDDARATTTFRG----- 99  
 QY 306 AASAVGAGVEDVYLAGALEAQNVAGEVYLEISDEEV-----DDGAGLPPASRR 353  
 DB 100 -----GERFFYLRLPAVDPLCYA--CILDSHSETVLNLEACVHGLEPCTPLPP--- 146  
 QY 354 RPVVGELWDDGPRRHERPTTTRIRHRKLSAYYRVAR-----PPVMTD--RLGVEV 404  
 DB 147 -PAPAE---ADG-----AARSVYARAARLATVAPPDQITPFVRLIRV 187  
 QY 405 FYGGRPAMSLVE-----RKVFILCSQNPPLADISHSLHSRKLGLVLLPKPD--DNNTGP 457  
 DB 188 FYFG-----SLVAETSQDQRRGRVLRHKQDPKP--GHECFYV-TAYKWLFPKQLDGLPTP 240  
 QY 458 GDNVLLAAVLASFPASGLVIVSLRSGIYVKNLCKSTVLYHGNP-----PKKF 504  
 DB 241 EQRETVCIELNGCEGVFLHGNELGMVVDNRTRTVACAGNDAEGNHAQRAVRSSVRSQI 300  
 QY 505 GVICGLSSRAVLDFVNAQYRIQHE-HIKTTVFIGDPTSAQFDMVPLVILKRLRSV 563  
 DB 301 FYVMGLLR-----LARSVPDGTVPFNSAVTLVGLGRPGSKR-PQVPVTL-----V 346  
 QY 564 TCDD 567  
 DB 347 ICQD 350  
 RESULT 14  
 Q9BIA3 PRELIMINARY; PRT; 947 AA.  
 ID Q9BIA3  
 AC Q9BIA3;

Search completed: June 9, 2004, 09:02:37  
Job time : 48 BECS